DOCUMENT RESUME

ED 412 216 TM 026 131

TITLE	Preschool Assessment Instruments.
INSTITUTION	Nevada State Dept. of Education, Carson City.
PUB DATE	1994-00-00
NOTE	134p.
PUB TYPE	Reference Materials - Bibliographies (131)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Criterion Referenced Tests; *Diagnostic Tests; Disabilities; Educational Assessment; Evaluation Methods; Identification; Norm Referenced Tests; *Preschool Children; Preschool Education; *Screening Tests; *Special Education; Technical Assistance; *Test Use; Testing Problems
IDENTIFIERS	*Nevada

ABSTRACT

This directory is intended to provide technical assistance for Nevada school districts and other evaluators of preschool children in selecting appropriate instruments to use in the multidisciplinary process for special education eligibility. It is not a list of approved tests, and tests not listed in this directory may be appropriate for assessment. It is the responsibility of parents and professionals working together to develop a picture of the child's abilities, and these instruments may be helpful. In all, 56 tests are listed. Each test description describes the author, publisher, publication information, the range of children tested, and details about test administration and use. A "Discussion/Concerns" section mentions additional detail about the test and its use. Charts also divided the tests into groups as: (1) norm-referenced; (2) criterion-referenced; and (3) measures that can be adapted for use with parents. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 216

PRESCHOOL ASSESSMENT INSTRUMENTS

NEVADA DEPARTMENT OF EDUCATION
FALL 1994

TM 026 131

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

NEVADA DEPARTMENT OF EDUCATION

Special Education Branch
400 W. King Street
Carson City, NV 89710

Nevada State Board of Education

Yvonne Shaw, President
Liliam L. Hickey, Vice President
Peggy Lear Bowen
Frank Brown
Carley Sullivan
William Hanlon
Patricia Krajcech
Carol Lenhart
Dr. Marianne Long
Steve Stallworth
Terry Garcia-Cahlan
Jenny Herndon, Student Representative

Nevada Department of Education

Eugene T. Paslov, Superintendent of Public Instruction
Mary L. Peterson, Deputy Superintendent of Instructional Services

Gloria Dopf, Director, Special Education
Ann Drendel-Haas, Consultant
Marsha Lakes, Consultant
Ann M. Marek, Consultant
Sharon Rogers, Consultant
Andrea Trow, Consultant

1994

INTRODUCTION

This document, Preschool Assessment Instruments, is intended to provide technical assistance to Nevada school districts and other evaluators of preschool children in selecting appropriate instruments to use in the multidisciplinary assessment process for special education eligibility. This resource should not be considered a listing of approved tests, as no endorsement or recommendation by the Nevada Department of Education is intended through the inclusion of any instrument in this document. An appropriate and comprehensive assessment for any given child requires that parents and professionals work together to determine the components which are necessary to provide a picture of the child's abilities. Specific regulations addressing eligibility determinations can be found in the Nevada Administrative Code, (NAC), Chapter 388.

While instruments included in this document frequently are used for the assessment of preschool children with disabilities, it is the responsibility of the multidisciplinary team to select the specific tests for each child. Tests which are not included in this document may be appropriate and can be utilized as part of the assessment or evaluation process. The Nevada Department of Education welcomes additional information on new and promising tests for preschool children suspected of having a disability.

Special acknowledgment is given to the Arizona Department of Education and Editor, Dr. Lynn Busenbark, who gave Nevada permission to reprint this material that was developed for AZ-TAS Themes and Issues in a series of topical papers on special education.

This guidance is provided as part of a larger technical assistance effort known as NECTAR. NECTAR stands for Nevada Early Childhood Technical Assistance Resources. It is the name given to the training and technical assistance activities provided by the Special Education Branch of the Nevada Department of Education. The acronym and logo were developed to represent the "cross pollination" efforts the Department conducts or sponsors to facilitate the sharing of information, ideas and techniques among early childhood special education programs in Nevada.

Norm-refe .d Measures

Name of Test	Page Number	Age Range: B	1	2	3	4	5	6	7	Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Developmental Areas: Motor	Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Technical* Data: Standardization	Reliability	Validity	Comments
Assessment Link between Phonology and Articulation (ALPHA)	13				X	X	X	X	X	N			D	P				X					X	X	X	
Battelle Developmental Inventory	15	X	X	X	X	X	X	X	X	N	C		D	P	X	X	X		X		X		X	X	X	
Bayley Scales of Infant Development	17	X	X	X						N			D		X	X	X			X			X	X	X	
Boehm Test of Basic Concepts	19					X	X	X	X	N		S		P			X						X	X	X	
Bracken Basic Concept Scale	21			X	X	X	X	X	X	N		S	D	P		X	X						X	X	X	
Burks' Behavior Rating Scales	23				X	X	X	X		N			D								X		X	X	X	
California Preschool Social Competency Scale	25				X	X	X	X		N			D						X		X		X	X	X	Normed-does not yield a standard deviation
Carrow Elicited Language Inventory-Preschool (CELI)	33				X	X	X	X	X	N			D	P			X						X	X	X	Normed-does not yield a standard deviation
Childhood Autism Rating Scale (CARS)	35			X	X	X	X	X	X	N			D	P		X	X		X		X		X	X	X	Normed-does not yield a standard deviation
Child Behavior Checklist (CBC)	37			X	X	X	X	X	X	N			D								X		X	X	X	
Cognitive Abilities Scale (CAS)	39			X	X					N			D			X	X			X		X	X	X	X	
Developmental Profile II (DPII)	43	X	X	X	X	X	X	X	X	N			D		X	X	X		X		X		X	X	X	Normed-does not yield a standard deviation
Developmental Test of Visual-Motor Integration (VMI) — 3rd Revision	47			X	X	X	X	X	X	N			D		X					X			X	X	X	
Differential Abilities Scale (DAS)	51			X	X	X	X	X	X	N			D	P		X	X			X			X	X	X	
Expressive One Word Picture Vocab. Test - Revised (EOWPVT-R)	55			X	X	X	X	X	X	N			D				X						X	X	X	
Kaufman Assessment Battery for Children (K-ABC)	65			X	X	X	X	X	X	N			D			X				X			X	X	X	

*Technical Data information available in manual.

BEST COPY AVAILABLE

gab093 hms (5/93) (Spec Ed 5156)

Norm-referenced Measures (continued)

Name of Test	Page Number	Age Range								Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Developmental Areas								Technical* Data			Comments
		B	1	2	3	4	5	6	7						Motor	Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Standardization	Reliability	Validity	
Learning Accomplishment Profile-D (LAP-D) Standardized	69			X	X	X	X			N			D	P	X	X	X						X	X	X	
Leiter International Performance Scale	71			X	X	X	X	X	X	N			D			X				X			X	X	X	
McCarthy Scales of Children's Abilities	73			X	X	X	X	X	X	N			D		X	X				X			X	X	X	
Motor-Free Visual Perception Tests (MVPT)	81					X	X	X	X	N			D							X			X	X	X	Normed-does not yield a standard deviation
Peabody Developmental Motor Scales (PDMS)	87	X	X	X	X	X	X	X	X	N			D	P	X								X	X	X	
Peabody Picture Vocabulary Tests (PPVT-R)	89			X	X	X	X	X	X	N		S	D				X						X	X	X	
Personality Inventory for Children (PIC)	91				X	X	X	X	X	N			D								X		X	X	X	
Photo Articulation Test (PAT)	93				X	X	X	X	X	N			D					X					X	X	X	
Preschool Language Scale - 3 (PLS-3)	97	X	X	X	X	X	X	X		N		S	D	P			X						X	X	X	
Scales of Independent Behavior (SIB)	99	X	X	X	X	X	X	X	X	N			D	P	X		X		X		X		X	X	X	
Sequenced Inventory of Communication Development-Revised (SICD-R)	103	X	X	X	X					N			D				X						X	X	X	
Social Skills Rating System (SSRS)	107				X	X	X	X	X	N		S		P							X		X	X	X	
Stanford-Binet Intelligence Scale	109			X	X	X	X	X	X	N			D			X							X	X	X	
Structured Photographic Expressive Language Test - Preschool (SPELT-P)	111				X	X	X			N			D				X						X	X	X	
Structured Photographic Expressive Language Test - II (SPELT-II)	113					X	X	X	X	N			D				X						X	X	X	
Tests for Auditory Comprehension of Language-Revised (TACL-R)	115				X	X	X	X		N		S	D	P			X						X	X	X	

*Tech[illegible]ata information available in manual.

Norm-reference asures (continued)

Name of Test	Page Number	Age Range								Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Developmental Areas								Technical* Data			Comments
		B	1	2	3	4	5	6	7						Motor	Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Standardization	Reliability	Validity	
Test of Early Language Development (TELD)	117				X	X	X	X	X	N			D	P			X						X	X	X	
Test of Early Socioemotional Development (TOESD)	119				X	X	X	X	X	N			D								X		X	X	X	
Test of Language Development - 2P (TOLD-2P)	121					X	X	X	X	N			D				X						X	X	X	
Vineland Adaptive Behavior Scales	125	X	X	X	X	X	X	X	X	N			D	P	X		X		X		X		X	X	X	
Wechsler Preschool and Primary Scale of Intelligence (WPPSI)	129				X	X	X	X	X	N			D			X	X			X			X	X	X	
Woodcock-Johnson Psychoeducational Battery	131			X	X	X	X	X	X	N			D			X			X		X		X	X	X	

*Technical Data information available in manual

BEST COPY AVAILABLE

Criterion-referenced Measures

Name of Test	Page Number	Age Range: B	1	2	3	4	5	6	7	Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Developmental Areas: Motor	Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Technical* Data: Standardization	Reliability	Validity	Comments
Arizona Behavior Analysis Criterion Utilization System (ABACUS)	11		X	X	X	X	X				C	S	D	P			X		X		X	X				
Battelle Development Inventory	13	X	X	X	X	X	X	X	X	N	C		D	P	X	X	X		X		X		X	X	X	
The Callier-Azusa Scale	27	X	X	X	X	X	X	X	X		C			P	X	X	X		X	X	X			X	X	Deaf-blind Severe/profound
Carolina Curriculums for Preschoolers	29	X	X	X	X	X	X	X	X		C		D	P	X	X	X		X		X					
Carolina Developmental Profile	31			X	X	X	X				C			P	X	X	X			X						
Cognitive Skills Assessment Battery	41					X	X	X			C			P								X	X			
Developmental Programming	45	X	X	X	X	X	X	X			C			P	X	X	X		X	X	X					
Diagnostic Inventory of Early Development (Brigance)	49	X	X	X	X	X	X	X	X		C			P	X	X	X	X	X	X	X	X		X	X	
The Early Learning Accomplishment Profile-Revised (E-LAP)	53	X	X	X	X						C			P	X	X	X		X		X					
The Fisher-Logemann Test of Articulation Competence	57				X	X	X	X	X		C	S		P				X					X			
Gesell Developmental Schedules-Revised	59	X	X	X	X						C	S	D		X		X		X	X	X	X	X	X	X	
Goldman-Fristoe Test of Articulation	61			X	X	X	X	X			C		D					X								
Hawaii Early Learning Profile (HELP)	63	X	X	X	X	X	X				C			P	X	X	X		X		X					
Milani-Comparetti Developmental Scale	77	X	X								C		D	P	X											Cerebral Palsy
A Motor Development Checklist	79	X	X	X							C			P	X											
Ordinal Scales of Psychological Development	83	X	X								C			P		X										Useful for the assessment of older severely disabled children.

*Tec[illegible] Da[illegible] [illegible]rmation available in manual

Criterion-referenc. measures (continued)

Name of Test	Page Number	Age Range								Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Motor	Developmental Areas								Technical* Data			Comments
		B	1	2	3	4	5	6	7							Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Standardization	Reliability	Validity	
Portage Guide to Early Education-Revised	95	X	X	X	X	X	X				C			P	X	X	X		X		X					
Skills Inventory–Oregon Project	105	X	X	X	X	X	X	X			C			P	X	X	X		X		X					Visually impaired only
Uniform Performance Assessment System (UPAS)	123	X	X	X	X	X	X	X			C			P	X		X		X		X			X		

13

*Technical Data information available in manual

BEST COPY AVAILABLE

Measures Which Can Be Adapted For Use With Parents

Name of Test	Page Number	Age Range								Norm-Referenced	Criterion-Referenced	Screening	Diagnostic	Programming	Developmental Areas									Technical* Data			Comments
		B	1	2	3	4	5	6	7						Motor	Cognition	Language	Speech	Self-Help	Perception	Social-Behavioral	Kdgn-Readiness	Standardization	Reliability	Validity	
Bayley Scales of Infant Development	17	X	X	X						N			D		X	X	X			X			X	X	X	
Burks' Behavior Rating Scales	23				X	X	X	X		N			D								X		X	X	X	
The Callier-Azusa Scale	27	X	X	X	X	X	X	X	X		C			P	X	X	X		X	X	X			X	X	Deaf-blind Severe/profound
Developmental Profile II (DP-II)	43	X	X	X	X	X	X	X	X	N			D		X	X	X		X	X	X		X	X	X	
Diagnostic Inventory of Early Development (Brigance)	49	X	X	X	X	X	X	X			C			P	X	X	X		X	X	X	X				
Hawaii Early Learning Profile (HELP)	63	X	X	X	X						C			P	X	X	X		X		X					
Portage Guide to Early Education-Revised	95	X	X	X	X	X	X				C			P	X	X	X		X		X					
Preschool Language Scale - 3 (PLS-3)	97	X	X	X	X	X	X	X		N		S	D	P			X						X	X	X	
Social Skills Rating System (SSRS)	107				X	X	X	X	X	N		S		P							X		X	X	X	
Uniform Performance Assessment System (UPAS)	123	X	X	X	X	X	X	X			C			P	X		X		X		X		X	X		
Vineland Adaptive Behavior Scale	125	X	X	X	X	X	X	X	X	N			D	P	X		X		X		X		X	X	X	

15

16

*Te[illegible] Dr[illegible]rmation available in manual

gab09[illegible] (Spec Ed 5156)

BEST COPY AVAILABLE

Arizona Behavior Analysis Criterion Utilization System (ABACUS)

Author(s):	Jeanne McCarthy, Kathryn Lund, Candace Bos, Jean Glattke, and Shari Vaughn
Publisher:	Love Publishing Co.
Address:	1777 South Bellaire Street, Denver, CO 80222
Copyright Date:	1981
Price:	$195 (Complete Kit) $29.95 (Scoring Booklets—20)
Purpose:	To assess young children for instructional purposes and to establish the child's present level of performance in each of five areas to include body management, self-help, communication, pre-academics, and socialization.
Description:	The examiner presents tasks for the child as presented in the individual record form for each area, or observes the child at natural play, in order to gather information.
Range of Children:	The ABACUS can be used to assess children from 18 months to 5½ years.
Testing Time:	It takes approximately 60 to 90 minutes to administer. Pass I or Pass II (Pre-Abacus screening scale) can be administered in 20 to 30 minutes.
Scoring:	Raw scores are based on the number of tasks a child can do, as a percent of total tasks.
Examiner:	Professional
Standardization:	This instrument is not normed.
Reliability:	No reliability studies are reported.
Validity:	No validity studies are reported.
Discussion/Concerns:	The ABACUS is a comprehensive profile which covers several developmental areas. Its foundation encompasses behavioral, developmental, and cognitive theories. The ABACUS includes a complete system of screening, assessment, IEP data management, curriculum and instruction, family involvement and teacher evaluations. It is designed to evaluate young handicapped children and is best used to formulate specific instructional objectives.

BEST COPY AVAILABLE

References:

McCarthy, J., Lund, K., Bos, C., Glattke, J., Vaughn, S. (1981). *Arizona Behavior Analysis Criterion Utilization System (ABACUS).* Denver, CO: Love Publishing Co.

West Virginia Dept. of Education. (1988). *Assessment of Preschool Handicapped Children.*

Assessment Link Between Phonology and Articulation (ALPHA)

Author(s):	Robert J. Lowe
Publisher:	LinguiSystems, Inc.
Address:	716 17th St., Moline, IL 61265 (309) 762-5112
Copyright Date:	1986
Price:	Contact Publisher

Purpose: The ALPHA is designed to provide a link between traditional articulation testing and phonological assessment. It is a delayed sentence imitation test that assesses a subject's phonetic inventory and a subject's deviant use of phonological processes.

Description: The ALPHA is an individually administered test which consists of 50 items. Each of 50 target words is embedded in a short sentence. Following the examiner's sentence model and presentation of a stimulus picture, the subject repeats the stimulus sentence. As the child is responding, the examiner phonetically transcribes the target word produced by the child. The examiner assesses the target words for articulation errors or analyzes the sound change errors for the presence of phonological process errors.

Range of Children: The ALPHA can be administered to children ages 0-3 to 8-11.

Testing Time: This test can be administered in approximately 10 to 15 minutes.

Scoring: Errors are transcribed onto an answer sheet which consists of two sections: phoneme error section and phonological processes section. Raw scores can be converted into percent of occurrence and compared to the mean percent of occurrence of the child's age group. Percents of occurrence are derived by dividing raw scores attainable for each process by the number of instances that the process could occur in the ALPHA test words. The raw scores of total processes can be converted into percentile ranks and standard scores.

Examiner: A professional who has experience in transcribing deviant articulations—usually a speech and language clinician.

Standardization: The norm group was based on a sample of 1,310 subjects who were randomly selected from approximately 13 cities in the Midwest. The sample was representative of the population with respect to sex and age. School-aged children were selected from schools, and preschool children were selected from day-care centers and preschool programs.

The sample was not restricted to any specific economic, intellectual, or racial requirements, but the sampling procedures were designed to ensure adequate representation of minority populations in accordance with the 1980 national census figures. The standardization sample was limited to normal children.

Reliability: Test-retest reliability of the ALPHA yielded correlation coefficients of .80 to .99 with a mean of .97. The interscorer reliability coefficient was .93. Internal consistency was not discussed in the manual.

Validity: Construct validity was reportedly established based on the following: (1) evidence of differences between normal speaking and articulation disordered groups, (2) changes in the phonological processes over age, (3) intercorrelations between phonological processes, and (4) phonological process correlations with scores of total processes. Content validity and concurrent validity were not discussed in the manual.

Discussion/Concerns: The ALPHA is a unique test as it attempts to link articulation with phonology. Its psychometric properties are somewhat weak in that the standardization sample is not adequately representative of the U.S. population and only includes normal children. Consequently, children who have articulation difficulties cannot be compared accurately to the norm group, which may result in misdiagnosis or overcategorization. Reliability of the ALPHA appears to be adequate while validity has not been sufficiently established. Overall, the ALPHA may provide valuable information for planning purposes and could be used as a screening instrument or as a supplemental tool to other instruments.

References:

Lowe, R. J. (1986). *Examiner's Manual: Assessment Link Between Phonology and Articulation*. Moline, IL: LinguiSystems.

Battelle Developmental Inventory (BDI)

Author(s):	Jean Newborg, John R. Stock, Linda Wnek, John Guidubaldi, & John Svinicki
Publisher:	DLM Teaching Resources
Address:	One DLM Park, Allen, TX 75002
Copyright Date:	1984
Price:	$199

Purpose: To identify the developmental strengths and weaknesses of handicapped and nonhandicapped children in infant, preschool, and primary programs.

Description: The BDI is a standardized, individually administered assessment battery of key developmental skills. Data are collected through presentation of a structured test format, an interview with parents and teachers, and observations of the child in the natural setting. The BDI consists of a screening component and a full battery. The full battery consists of 341 items grouped into five domains: personal-social, adaptive, motor, communication, and cognitive. Each domain is contained in a separate test booklet. Each item is presented in a standard format that specifies the behavior to be assessed, the materials needed for testing, the procedures for administering the items, and the scoring criteria. The screening test consists of 96 of the 341 items. Not all materials are included in the kit and must be furnished by the examiner.

Range of Children: Birth to 8 years

Testing Time: 1 to 1½ hours (BDI)
10 to 30 minutes (Screening Test)

Scoring: Each item is scored on a three-point scale (0,1,2) which emphasizes emerging versus fully-acquired developmental skills according to specified criteria. A score can be obtained in each domain as well as a total performance score for the full BDI. Individual domain scores are totaled to obtain the total raw score. The authors caution against the use of age-equivalent scores. Thus, raw scores can be converted to a variety of derived scores (e.g., percentiles, Z-scores, T-scores, deviation quotients, and normal curve equivalents).

Examiner: Teachers, diagnosticians, and multidisciplinary teams.

Standardization: The normative sample was stratified, according to 1980 census information, by geographical regions and subregions, race, and sex. Eight hundred children were tested in 24 states, approximately 50 of each gender at each age level.

Reliability: Reliability is adequate for the BDI. A test-retest reliability coefficient .99 was reported with regard to a sample of 183 subjects tested over a four-week retest time. Interrater reliability coefficients for 148 subjects was .99. In addition, the reported standard error of measure is quite small.

Validity: Validity information is limited. The authors report that content validity was ensured in the lengthy test development process.

Discussion/Concerns: Overall the BDI appears to be a useful standardized instrument for developmental assessment of young handicapped and nonhandicapped children. The manual is informative and comprehensive, making scoring and administration of the test manageable. The reliability and standardization are adequate but further validity studies are needed. There are additional concerns about the administration of the test. Reliability could be threatened if precautions stated by the authors are not taken when a multidisciplinary team administers the various assessments. There is concern with regard to the number of items. According to Paget (1989), in some domains there is only one item at some of the age levels. When this occurs it would not be advisable to use these performances to support the need for special education services.

References:

Newborg, J., Stock, J., & Wnek, L. (1984). *Battelle Developmental Inventory*. Allen, TX: DLM Teaching Resources.

Paget, K. D. (1989). Review of the Battelle Developmental Inventory. In J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (Vol I, pp 70-73). Lincoln: University of Nebraska Press.

Bayley Scales of Infant Development (BSID)

Author(s):	Nancy Bayley
Publisher:	Psychology Corporation
Address:	Order Service Center, Box 839954, San Antonio, Texas, 78283-3954
Copyright Date:	1969
Price:	$550 per kit $12 for supplemental manual

Purpose: The scales assess developmental status in infants and toddlers.

Description: The instrument consists of two scales. The Mental Scale (163 items) measures (1) sensory-perceptual acuities and discrimination; (2) early acquisition of object constancy and memory, learning, and problem-solving ability; (3) vocalizations and the beginning of verbal communication; and (4) early evidence of the ability to form generalizations and classifications. The Motor Scale (81 items) measures the degree of and control of the body, coordination of the large muscles, and finer manipulatory skills of the hands and fingers.

Range of Children: The scales can be administered to children ranging in age from 2 months to 30 months.

Testing Time: Average testing time is 45 minutes to 75 minutes.

Scoring: The test items are arranged by age level on the scoring forms, and each item passed is given one point. Raw scores can be converted to standard scores.

Examiner: Professional, usually a psychologist. The mother or primary caregiver is usually present during test administration.

Standardization: The scales were standardized on a sample of 1,262 infants and children ranging in age from 2 to 30 months. The sample was selected to be representative of the U.S. population in terms of major geographic areas and residence (urban, rural) and was controlled for sex, race, and education level of head of household. Only "normal" children living at home were included in the sample. Infants from urban areas and higher parent educational levels were overrepresented in the norm group. (Roszkowski, 1989). Furthermore, these norms were developed in the mid-1960s and are considered outdated.

Reliability: Reliability coefficients reported for internal consistency for 14 age groups range from .62 to .93 (median of .84) on the Motor Scale. Reliability coefficients for the Mental Scale ranged from .81 to .93 (median of .88). Test-retest and interrater reliability were based on only one age group and only moderate reliability was found.

Several researchers have also taken on the task of assessing the Bayley's reliability and have collectively found that the "reliability of the BSID equals or is better than that of the other tests of its genre" (Roszkowski, 1989).

Validity: Concurrent validity between the BSID and the Stanford-Binet: Form L-M was .57 using one age group (two-year-olds). The manual does not discuss content, construct, or predictive validity. However, others have found some indications of adequate construct validity, especially among high-risk infants. Roszkowski (1989) also states that the "BSID is not a good predictor of future intelligence with normal infants, (but) . . . improves for impaired infants."

Discussion/Concerns: Reviewers and users of this instrument have found it somewhat cumbersome to administer. The test itself is colorful and durable, and scoring procedures are easy to follow. In general, the BSID is considered well-standardized with adequate psychometric characteristics. However, a correlation of .57 is considered inadequate concurrent validity. The author of the BSID states that the value of these scales in clinical practice does not lie in predicting a child's later abilities but in establishing his/her current developmental status in relation to others of the same age. A strength of the BSID is that the mother or primary caregiver can report the occurrence of test behaviors which are not observed by the test giver. A modified version of the BSID which is designed to assess cognitive development in handicapped infants (two to 30 months) by circumventing the effects of their physical limitations is also available.

References:

Bayley, N. (1969). *Bayley Scales of Infant Development Manual.* San Antonio, TX: Psychological Corporation.

Hoffman, H. (1974). *The Bayley Scales of Infant Development: Modifications for Youngsters with Handicapped Conditions.* Commack, NY: Suffolk Rehabilitation Center.

Rhodes, L., Bayley, N., & Yow, B.C. (1984). *Supplement to the Manual for the Bayley Scale of Infant Development.* San Antonio, TX: Psychological Corporation.

Roszkowski, M. J. (1989). Review of the Bayley Scales of Infant Development. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook.* Lincoln: The University of Nebraska Press.

Boehm Test of Basic Concepts-Preschool Version

Author(s)	Ann E. Boehm
Publisher:	Psychological Corporation
Address:	555 Academic Court, San Antonio, TX 78024-2498
Copyright:	1986
Price:	$80

Purpose: The Boehm Test of Basic Concepts-Preschool Version is designed to assess young children's mastery of basic relational concepts considered necessary for achievement in the beginning years of school.

Description: The Boehm-Preschool measures children's understanding of 26 basic concepts. It consists of 52 items (two items per concept) that are individually administered. The child is shown a picture from the picture book and is asked to respond to the examiner's question by pointing to an appropriate portion of the picture that shows the requested concept.

Range of Children: 3-5 years

Testing Time: 10-15 minutes. For children with limited attention spans, the author reports that the test may be administered over two testing sessions.

Scoring: Items are scored during the testing session. The examiner enters a "1" in the Score Column if the child answered the question correctly and a "O" if the child answered incorrectly. Each concept consists of two items so the child essentially has two opportunities per concept to demonstrate understanding. Raw scores for each of the 26 concepts are calculated separately and are then summed to yield a total score. Raw scores are then converted to T-scores and percentile ranks.

Examiner: Teachers, paraprofessionals, psychoeducational specialists, special educators, speech/language therapists, and others concerned with early childhood development.

Standardization: The standardization sample seems to have been adequately constructed. It consisted of 433 children from 17 states at over 35 different test sites within four major regions of the United States. The children were divided in five age levels in one-half year increments between 3 and 5 years of age. There were equal numbers of males and females. The sample was stratified with respect to race, region, and socioeconomic status which were all based on 1980 Census data.

Reliability: Two types of reliability were calculated: internal consistency and test-retest reliability. Internal consistency coefficients were calculated using coefficient alpha and split-half coefficients. Alpha coefficients across the five age levels ranged from .85 to .91 with the average being .88, and split-half coefficients ranged from .80 to .87 with the average being .85.

Test-retest reliability was also adequate. A test-retest analysis was conducted on a sample of 78 children, ages 3½ and 4½, with a retest interval of 7-10 days. The reliability coefficients were .94 and .87 for 3½- and 4½-year-old children respectively.

Validity:

Two types of validity were reported: content and concurrent. The author states, "content validity of the Boehm-Preschool has been inherent since its inception because items selected are represented in curriculum materials that are used frequently by preschool teachers."

A concurrent validity analysis was conducted with 29 children, mean age of 3-10, with both the Boehm-Preschool and the PPVT-R, Form L. The two instruments have a correlation of .63 which Boehm states demonstrates "a substantial relationship between a child's ability to understand basic concepts and a child's receptive vocabulary."

Discussion/Concerns:

Overall, the Boehm seems to be an adequate test for its stated purpose. It is basically psychometrically sound, although reliabilities are lower than desired and additional validity studies should be conducted. The test is easy to administer and score. The child is given two chances per concept to determine whether the concept is understood by the child or is just emerging. While too narrow a construct to stand alone, assessment of concept knowledge is an important component of a language assessment.

References:

Boehm, A. E. (1986). *Boehm Test of Basic Concepts Preschool Version*. San Antonio, TX: Psychological Corporation.

Bracken Basic Concept Scale (BBCS)

Author(s):	Bruce A. Bracken
Publisher:	Psychological Corporation
Address:	555 Academic Court, San Antonio, TX 78204-2498
Copyright Date:	1984
Price:	$140

Purpose: The BBCS was designed to assess a child's knowledge of basic concepts.

Description: The BBCS consists of three norm-referenced instruments: The Diagnostic Scale (DS) and two parallel forms of a group-administered 30-item Screening Test.

The DS measures 258 concepts divided into 11 subtests including colors, letter identification, numbers/counting, comparisons, shapes, direction/position, social/emotional, size, texture/material, quantity, and time/sequence. The test format requires the examiner to read a statement; the child responds by pointing to a picture. No verbal responses are required. The DS yields three scores: (1) the first five subtest scores are combined to form a School Readiness Composite standard score; (2) the remaining six subtests have individual scores; and 3) a composite score is calculated from the standard scores for all 11 subtests.

The 30-item Screening Test was designed to identify kindergarten and first-grade children who may benefit from a more intensive diagnostic assessment of conceptual knowledge.

Range of Children: The DS is appropriate for children between the ages of 2 years, 6 months and 7 years, 11 months.

Standardization: The author indicates that the standardization sample for the DS, which consisted of 1,109 children between the ages of 2½ and 8 years, was representative of the 1980 census figures.

Reliability: According to reviewers, test-retest reliability of the DS subtests is inadequate for diagnostic purposes. However, it is sufficiently high for screening purposes. The sample was rather limited; 27 children took the test 14 days apart. The coefficients ranged from .67 (Size subtest) to .98. Internal consistency coefficients for the DS ranged from .47 to .96. According to Ysseldyke (1986), the coefficients for the Size, Social/Emotional, Texture/Material, and Time/Sequence subtests are too low for diagnostic purposes at age 6.

Validity: The author reports content validity and concurrent validity evidence for the BBCS. Bracken claims that content validity is good because the items on the scale are often used in preschool and primary tests typically given to young children. Eight validity studies are reported in the manual. Concurrent validity is adequate, from .68 to .88, with regard to other measures of basic concept (e.g., Boehm Test of Basic Concepts and the Peabody Picture Vocabulary Test-Revised). Since the author fails to report predictive validity, we do not know whether a low BBCS score is predictive of any academic deficits (Ysseldyke, 1989).

Discussion/Concerns: For the most part, the test is technically adequate; however, there is some concern about its adequacy as a diagnostic test. It is a useful instrument for identifying concepts that students do and do not know and for planning interventions to teach these concepts. Those wanting to make screening decisions should use the BBCS Diagnostic Scale, as it has better statistical data than the more limited Screening Test.

References:

Bracken, B. (1984). *Bracken Basic Concepts Scale*. San Antonio TX: Psychological Corporation.

Turco, T. L. (1989). Review of Bracken Basic Concept Scale. In J. Close-Conley & J. Kramer (Eds.), *Tenth Mental Measurements Yearbook* (Vol. I, pp. 102-104). Lincoln: University of Nebraska Press.

Ysseldyke, J. E. (1989). Review of Bracken Basic Concept Scale. In J. Close-Conley & J. Kramer (Eds.), *Tenth Mental Measurements Yearbook* (Vol. I, pp. (104-105). Lincoln: University of Nebraska Press.

Burks' Behavior Rating Scales (BBRS) (Preschool and Kindergarten Edition)

Author(s):	Harold L. Burks
Publisher:	Western Psychological Services
Address:	Publishers and Distributors, 12031 Wilshire Boulevard, Los Angeles, CA 90025
Copyright Date:	1977
Price:	$33.50—Kit (Preschool/Kindergarten)

Purpose: The Burks' Behavior Rating Scale, Preschool and Kindergarten Edition specifically attempts to gauge the severity of negative symptoms as seen by outside persons: ordinarily teachers or parents.

Description: The Scales consist of 105 items which are rated by respondents who make a quantitative judgement by determining the degree to which each identified behavior is seen in the child being rated. The items have been clustered into 18 groupings. The subscales have been named according to the type of behavior shown. Raters judge each behavior described according to a five-point scale based on the frequency of the behavior. The scales attempt to measure the degree of severity of behavior observed over a considerable period of time, *not* conduct observed at the present moment.

Range of Children: The Preschool and Kindergarten Edition has been designed for children ages 3 through 6.

Testing Time: No information provided.

Scoring: The total scores for each subscale are recorded on the Profile Sheet. Each score is then plotted on a horizontal number line. The number line is divided into three sections: Not Significant, Significant, Very Significant. By placing an X at a point along each category continuum where score and number coincide, the one who scores the BBRS is able to determine the degree to which a particular subscale represents problem behavior for the child.

Examiner: Respondents can be anyone who knows the child well. The author recommends that the results be scored by a professional person (other than the rater).

Standardization: The research sample consisted of a group of 3- and 4-year-old preschool children and a group of kindergarten students from San Bernardino County, CA. In addition, 337 kindergarten children (184 boys, 153 girls) were rated by 31 teachers in four school districts located in California. No breakdown according to race or socioeconomic level is provided.

Reliability: Test-retest reliability was established by having 84 kindergarten children rated and rerated within a period of ten days by their teachers. Correlation coefficients ranged from .74 to .96.

Validity: The 105 scale items were chosen from a large pool of items originally constructed by the author. Many of these items or variations of them had been employed in a previous checklist devised by the writer and validated against quantifiable criteria. The categories were formed after factor analysis of the 110 items found in the scale for older children. Nineteen groupings were discovered and 18 were retained for the Pre-School and Kindergarten Edition.

Factorial validity studies were carried out by obtaining ratings from 127 preschool children. Results indicated that three patterns of probable behavior emerged. Similar factors emerged from teacher ratings of kindergarten children.

Discussion/Concerns: The Burks' Behavior Rating Scale provides a system for rating a child according to 105 behaviors by adults who know the child well. Responses made by adults who know the child in different environments can be compared. BBRS items are clearly stated and the rating scale can be used with minimum difficulty.

Martin (1991), in a review of this instrument, notes that interpretive guidelines are based on very poor technical details. Also, the preschool version includes items from the Burks' Behavior Rating Scales for school-aged children which do not adequately assess the developmental levels of young children. A strength of this instrument is that it samples a wide range of problem behaviors. The Burks' Preschool and Kindergarten Edition should not be used alone as a diagnostic instrument, but may be useful as part of an assessment battery.

References:

Burks, H. F. (1977). *Burks' Behavior Rating Scales Manual.* Los Angeles: Western Psychological Services.

Martin, R. P. (1991). Assessment of Social and Emotional Behavior. In B. A. Bracken (Ed.), *The Psycho-educational Assessment of Preschool Children* (pp. 450-464). Needham, MA: Allyn & Bacon.

The California Preschool Social Competency Scale

Author(s):	S. Levine, F. F. Elzey, M. Lewis
Publisher:	Consulting Psychologist Press, Inc.
Address:	Box 10096, Palo Alto, CA 94303
Copyright Date:	1969
Price:	$28 per Kit

Purpose: The Scale is designed to measure the adequacy of interpersonal behavior and degree of social responsibility in children ages 2-5. The behaviors included are situational in nature. They were selected in terms of common cultural expectations to represent basic competencies to be developed in the process of socialization.

Description: The Scale contains 30 items. Each item consists of four descriptive statements, given in behavioral terms, and representing varying degrees of competency. The examiner must have had considerable opportunity to observe the child in a variety of situations prior to completion of the scale.

Range of Children: The Scale was designed for use with children ranging in age from 2½ to 5½ years of age.

Testing Time: Approximately 15 minutes to complete rating scale.

Scoring: Each item contains four descriptive statements representing varying degrees of competency. The examiner circles the number of the option that is most characteristic of the child. All circled numbers are added to obtain a total score. Raw scores are then converted to percentiles. Additional rating instructions are provided to aid in the determination of the appropriate level for certain items.

Examiner: Professional. The Scale was explicitly developed to be used by teachers within the context of a preschool program, but can be utilized by other professionals who are familiar both with scale items and with the child's functioning.

Standardization: The norming sample approximates the proportion of preschool children in the major urban centers for each geographic region of the United States. The norming sample was based on ratings of 800 children. The norms for the scale were established by determining the percentile rank of the social competency raw scores for each chronological age by sex and occupational level or parents. The mean and standard deviation of the raw scores at each age level for each group were used for the computation of the norms. The norm group is small and outdated.

Reliability: To establish interrater reliability, studies were conducted in Texas, Minnesota, and California. Correlation coefficients ranged from .75 to .86. However, a small sample size was used (Lytton, 1978). Odd-even reliability coefficients ranged from .90 to .98. Internal consistency was found to be adequate.

Validity: A reviewer of the test, Calfee (1978), felt the test was constructed very subjectively and inappropriately reflected the values of preschool teachers. He believed test items were not structured well enough. Another reviewer (Prager, 1978) felt test items were well-constructed with 27 of the 30 items sampling preschool behaviors. Reviewers generally agree that concurrent, construct, and predictive validity have not been established for this assessment (Calfee, 1978; Lytton, 1978; Prager, 1978).

Discussion/Concerns: The California Preschool Social Competency Scale does not appear to have adequate psychometric properties. Its norms are outdated, and validity and reliability are considered inadequate. It is not a comprehensive assessment of social competency as it includes a limited item pool. It is not a good diagnostic device and would "fare much better if it were part of comprehensive battery of instruments" (Prager, 1978).

References:

Calfee, R. C. (1978). Review of the California Preschool Social Competency Scale. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*. Highland Parks, NJ: The Gryphon Press.

Levine, S., Elzey, L. and Lewis, M. (1969). *California Preschool Social Competency Scale*. Palo Alto, CA: Consulting Psychologists Press.

Lytton, H. (1978). Review of the California Preschool Social Competency Scale. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*. Highland Parks, NJ: The Gryphon Press.

Prager, B. B. (1978). Review of the California Preschool Social Competency Scale. In O. K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*. Highland Parks, NJ: The Gryphon Press.

The Callier-Azusa Scale-H

Author(s):	Robert Stillman & Christy Battle
Publisher:	The University of Texas at Dallas Callier Center for Communication Disorders
Address:	1966 Inwood Road, Dallas, TX 75235
Copyright Date:	1985
Price:	$12

Purpose: The Callier-Azusa Scale-H is an assessment scale designed to offer the educator and clinician a comprehensive, developmentally based framework for viewing the communicative abilities of deaf-blind and severely/profoundly handicapped persons. It provides assessment information relevant to planning developmentally appropriate activities for the child.

Description: The Callier-Azusa is composed of a hierarchical progression of items within four developmental domains: Representational and Symbolic Abilities, Receptive Communication, Intentional Communication, and Reciprocity. Each of the four subscales is divided into sequential steps identified numerically. Items within a level are not listed sequentially. Examples which are representative of behaviors typical of deaf-blind and severely-profoundly handicapped persons are provided.

The administration of the Callier-Azusa Scale is based on observation of behaviors which typically occur in conjunction with classroom activities. The authors recommend that more accurate results are obtained if several individuals having close contact with the child evaluate the child on a consensus basis and that the child be observed for at least two weeks in the classroom prior to the completion of the scale.

Range of Children: Birth to 108 months (9 years)

Testing Time: Observational tool, therefore, no testing time is indicated.

Scoring: The examiner circles the number on each subscale column corresponding to the level at or below which all abilities have been demonstrated. Then, the examiner circles the letter(s) corresponding to items describing demonstrated higher level abilities. Approximate age equivalencies for the items are provided in the left column. These are only rough estimates and should be treated as such.

Examiner: Anyone familiar with the child's behavior.

Standardization: The selection of items and examples was guided by the authors' conception of underlying developmental processes and how these processes are revealed in behavior.

Reliability: Reliability data are available from the publisher.

Validity: Validity data are available from the publisher.

Discussion/Concerns: Unlike most assessment instruments, the Callier-Asuza Scale is not administered directly item-by-item. Instead the evaluator is to complete the instrument based upon behaviors which appear spontaneously and are appropriately generalized within the context of routine classroom activities. The manual recommends that several individuals familiar with the child (teachers, aides, parents, etc.) can complete the scale together. As one might expect in an observational tool, only general guidelines for administering and crediting items are provided. For some of the items, examples of behaviors which can be used to credit items are given.

Persons completing the scale are to determine the base step, i.e., the number of the item in a subscale at which all the lower behaviors (steps and/or substeps depending on the scale) are an integrated part of a child's repertoire. All behaviors above the base step which are an integrated part of the child's repertoire are then determined. The instrument does not provide corresponding guidelines for determining the ceiling level. No estimates of time needed to complete the instrument or how often the assessment should be administered are given. (Assessment of Preschool Handicapped Children, West Virginia Department of Education, Feb. 1988)

References:

Stillman, R. & Battle, C. (1985). *Callier-Azusa Scale (H) for the Assessment of Communication Abilities*. Dallas: University of Texas Callier Center For Communication Disorders.

The Carolina Curriculum for Infants and Toddlers with Special Needs (CCITSN)

The Carolina Curriculum for Preschoolers with Special Needs (CCPSN)

Author(s):	Nancy M. Johnson-Martin, Susan M. Attermeier, and Bonnie Hacker
Publisher:	Paul H. Brookes Publishing Co. Inc.
Address:	PO Box 10624, Baltimore, MD 21285-0624
Copyright Date:	1990, 1991
Price:	$40 (CCITSN), $36.50 (CCPSN)

Purpose:

Both the CCITSN and CCPSN are used to plan individualized educational programs. The goal of both curricula is to provide assistance to early intervention personnel, families of children with disabilities, and other caregivers as they strive to optimize children's interactions with their world and people in it.

Description:

The CCITSN is a revision of the Carolina Curriculum for Handicapped Infants at Risk (CCHIR). The CCPSN was revised first, and the authors felt that the CCHIR needed to be more congruent with regard to sequence and other changes made in the CCPSN. Other changes included more adaptations to accommodate particular handicapping conditions, significant revision of the communications section, and more attention to integrating intervention activities into the child's daily routine. In addition, the title of the CCHIR was changed to the CCITSN. The CCPSN picks up where the CCITSN leaves off.

Both have curricular sequences within the five developmental domains: Cognition, Communication, Social Adaptation, Fine Motor Skills, and Gross Motor Skills which are further divided into 26 separate teaching sequences. With each item teaching procedures, evaluation criteria, and intervention strategies are provided for youngsters with mild, moderate, severe, or multiple handicaps. Suggestions for integrating intervention activities into normal daily routines to make intervention a more natural part of the child's life and to ensure generalization are emphasized.

Both curriculum formats are similar so basic descriptions can essentially be applied to both. To plan the educational program, the child needs to be assessed to determine what skills the child has mastered and what skills need to be learned next. The Assessment Log on which all of the items in the curriculum are listed is used to determine which activities may be included in the intervention plan.

Range of Children:

Birth to 2 years (CCITSN), 2-5 years (CCPSN)

Testing Time: The authors report that assessment without prior observation may take up to two hours and assessments with prior observation may take no more than 30 minutes. The assessment may take place in one day or over several weeks.

Scoring: Items to be administered are listed on the Assessment Log. These are scored as passed, failed, or emerging according to criteria within the manual. Results are transferred to the Developmental Progress Chart which is a visual display of the child's progress throughout the curriculum sequences.

Examiner: The curriculum is used by paraprofessionals or professionals (educators, psychologists, day care workers, public health nurses, physical and occupational therapists, and speech/language specialists).

Reliability: No information is available.

Validity: No information is available.

Standardization: Items were chosen from lists of developmental skills from a variety of norm-referenced tests which were reviewed. Pertinent skills were incorporated (Bayley, 1969, Knoblock & Pasamanick, 1974). Skills were added which were defined by tests of development (The Ordinal Scales of Psychological Development Callier-Azusa Scale).

Discussion/Concerns: The manual is informative. In addition to the descriptions of procedures and scoring criteria, the manual lists many types of conditions that affect children having special needs, such as speech disorders, hyperactivity/attention deficit disorder, mental retardation, autism, visual and hearing impairments, etc. Also, sections providing descriptions of characteristics of the conditions, the effects in the classroom, classroom tips, and specialists who can help are provided. The manual also explains how to select items for the individual child's intervention program and developing and implementing the IEP.

The Carolina Curriculum is not a standardized instrument, but a compilation of skills thought to be relevant according to skills and items listed in other norm-referenced developmental skills tests and opinions of specialists in their fields.

References: Johnson-Martin, N.M., Attermeier, S.M., & Hacker, B. (1989). *The Carolina Curriculum for Preschoolers with Special Needs*. Baltimore: Paul H. Brookes Publishing.

Carolina Developmental Profile (CDP)

Author(s): David L. Lillie and Gloria L. Harbin
Publisher: Kaplan School Supply Corporation
Address: 600 Jonestown Road, Winston-Salem, NC 27103
Copyright Date: 1980
Price: $119 Profile Kit

Purpose: The CDP is an individually administered criterion-referenced test consisting of a test booklet with instructions included on the inside cover and a profile on the back cover. The CDP covers developmental abilities in six areas: Gross Motor, Fine Motor, Visual Perception, Reasoning, Receptive Language, and Expressive Language.

Description: "The Carolina Developmental Profile is designed to be used with the Developmental Task Instruction System. In this system, the goal is to increase the child's developmental abilities to the maximum level of proficiency to prepare him for the formal academic tasks he will face in the early elementary school years. The Profile is designed to assist the teacher in establishing long-range objectives to increase developmental abilities in six areas. The purpose of the checklist is not to compare or assess the child in terms of age normative data." (Manual)

Range of Children: 2 years-5 years

Testing Time: Testing time varies with the individual child.

Scoring: The items on the Carolina Developmental Profile are presented sequentially by area. A task number, a description of the task, and a developmental age are given for each item. If the child completes the task successfully, the examiner checks the "Can do" column; if not, places a mark in the "Cannot do" column.

The checklist should be given to a child in a large room and in several different sessions. The examiner starts with the items on the gross-motor checklist, selecting the age level at which she believes the child will have success. If the items are easily passed, the examiner may want to skip several age levels to find the child's base level of success. If all tasks at a particular age level are passed, it can be assumed that the child can do all tasks below that level in the specific developmental area.

The child's highest level of functioning is established by determining his/her Developmental Age Ceiling (DAC). The criterion for determining the DAC varies from one developmental area to another. At the end of each developmental section, list on the appropriate line the numbers of those tasks that the child cannot do below and at his/her Development Age Ceiling. The remainder of the long-range objectives should be taken from the next highest DAC level.

To summarize findings in all six areas of development on the front of the Profile, the examiner should first draw a line through all the tasks the child completed successfully. Next, indicate the long-range objectives for each developmental area by circling those task numbers on the number chart. Finally, write in the area indicated the priority objectives: those objectives that should receive the greatest amount of instruction time.

Examiner: Classroom personnel

Standardization: The items on the checklist were developed after a careful review of a variety of standardized developmental tests and scales. In these instruments, similar items are standardized and age ranges given. The Profile also contains age designations, but care should be taken not to apply these age ranges precisely. The purpose of this checklist is not to compare or assess the child in terms of age norms.

Reliability: Same as Standardization.

Validity: Same as Standardization.

Discussion/Concerns: The test is easily administered and scored. All visual materials necessary are included in the profile. However, the simple, black-and-white pictures are quite unattractive and not very stimulating for young children. Additional materials such as blocks, balls, scissors, etc., are not included and must be gathered before administration. Criteria for each item being successfully scored are included for easy reference. There are only five items in each age group for each area tested. This seems quite limited in the amount of information provided to you. Additionally, technical information is limited.

References: Lillie, D. L. *Early Childhood Education—An Individualized Approach to Developmental Instruction.* Science Research Associates.

Carrow Elicited Language Inventory (CELI)

Author(s):	Elizabeth Carrow-Woolfolk
Publisher:	DLM
Address:	One DLM Park, Allen, TX 75002
Copyright Date:	1974
Price:	$75 for complete test package

Purpose: According to the author, the CELI is a diagnostic test which provides a means for measuring a child's productive control of grammar. The instrument elicits imitations of a sequence of sentences that have been systematically developed to include basic sentence construction types and specific grammatical morphemes. "In addition to providing a means of identifying children with language problems, it can be used to determine which specific linguistic structures may be contributing to the child's inadequate linguistic performance" (Carrow, 1974).

Description: The CELI consists of 52 stimuli which include 51 sentences and one phrase. Sentences range from short and simple to lengthy and complex. Sentences were lengthened primarily by increasing the number of semantic relations, by phrase expansion, and by increasing the number of grammatical morphemes. Of the 51 sentences, 47 are in the active and four are in the passive voice. The test is administered by recording on tape the child's imitations of the stimulus sentence as presented by the examiner.

Range of Children: The CELI can be used to assess children from ages 3-0 to 7-11.

Testing Time: It takes approximately 45 minutes for administration, transcription, and scoring of the CELI.

Scoring: Scoring is completed by listening to the child's taped responses and transcribing errors onto a form resembling a matrix where the grammatical features are classified. Raw scores can be converted to mean scores, percentile ranks, and standard scores (stanines).

Examiner: Professional, usually speech clinician.

Standardization: Norms are based on a sample of 475 Caucasian children between the ages of 3-0 and 7-11 who were from middle socioeconomic-level homes and spoke standard American English. All the children were from Houston and were selected from day care centers and church schools. Children with apparent speech or language disorders were eliminated from the study.

BEST COPY AVAILABLE

Validity: Carrow reports that concurrent validity, using a sample of 20 children, was assessed by comparing the CELI and the Developmental Sentence Scoring test and was found to be .79. The author also indicates that the CELI effectively differentiates children without language delays from children with apparent language deficiencies. Also, test scores do improve with age and follow a developmental pattern.

Reliability: Test-retest reliability was assessed using 25 children; a correlation coefficient of .98 was obtained. Interrater reliability was assessed using two examiners and 10 randomly selected tapes. The correlation coefficient was .99. Another interrater reliability study involved 10 children who were previously diagnosed with a language disorder and found a correlation of .99 (Carrow, 1974).

Discussion/Concerns: Because Carrow considers the CELI to be a diagnostic assessment tool, there are concerns regarding its psychometric limitations. Content validity appears adequate while concurrent validity is considered inadequate. The standardization procedures are inadequate because the norm group was not representative of the population at large. The norm group did not include children with language problems or minority children which brings into question the acceptability of using this assessment tool with these populations. The author does state that the CELI may not be useful with children who misarticulate to the point of interfering with intelligibility, use severe jargon, and have severe echolalia. Reliability appears adequate. A strength of the test is that the child's responses are taped and can be reviewed at any time which aids in scoring accuracy. Also, administration and scoring appear straightforward. This test may be beneficial for assessing language problems in a very specific, formalized manner, however, for actual diagnostic purposes, the CELI should be used with caution.

References:

Carrow, E. (1974). *Carrow Elicited Language Inventory—Manual.* Austin, TX: DLM Teaching Resources.

Childhood Autism Rating Scale (CARS)

Author(s):	Eric Schopler, Robert J. Reichler, & Barbara Renner
Publisher:	Western Psychological Services
Address:	12031 Wilshire Blvd., Los Angeles, Ca. 90025
Copyright Date:	1971
Price:	$35 (25 rating scales and manual)

Purpose: The CARS helps to identify children with autism and to distinguish them from developmentally handicapped children who are not autistic. It also distinguishes between mild-to-moderate and severe autism.

Description: The CARS is typically administered at the end of the child's first diagnostic session (Schopler et al., 1980). After the child has been observed and relevant information from parent reports and/or other records has been examined, the CARS is completed. This instrument consists of 15 scales including the following: (1) impairment in human relationships, (2) imitation, (3) inappropriate affect, (4) bizarre use of body movement and persistence of stereotypes, (5) peculiarities in relating to nonhuman objects, (6) resistance to environmental change, (7) peculiarities of visual responsiveness, (8) peculiarities of auditory responsiveness, (9) near receptor responsiveness, (10) anxiety reaction, (11) verbal communication, (12) nonverbal communication, (13) activity level, (14) intellectual functioning, and (15) general impressions.

Range of Children: The CARS can be used with any child over 2 years of age.

Testing Time: Testing time is highly variable.

Scoring: Each of the scales is scored on a continuum from normal to severely abnormal. Using a seven-point scale, the degree to which the child's behavior deviates from that of same-age peers is recorded. The total CARS score for each child has a possible range from 15 to 60. Individuals with scores of less than 30 are not considered autistic. Within the autistic range, two groups exist: severely autistic and mild-to-moderately autistic.

Examiner: Professional such as physicians, school psychologists, and special educators.

Standardization: The standardization sample included 537 children who were assessed using the CARS over a ten-year span. The children were assessed through TEACCH located in Chapel Hill, North Carolina. Seventy-five percent of the subjects were male and 25 percent female. Both boys and girls had approximately the same age distribution: 55 percent were below six years old and 11 percent were 10 years or older. The socioeconomic status of the majority of children was low.

Approximately three-quarters of the sample was made up of Caucasian children with the rest being African-American. Seventy percent of the population had I.Q.'s below 70. According to the Western Psychological Services Catalogue, as of 1991, over 15,000 cases have been assessed by the developers of the CARS.

Reliability: The internal consistency coefficient of the CARS is reported to be .94. Interrater reliability, using 280 cases, was found to range from .62 to .93 with a mean of .71 (Schopler et al., 1980).

Validity: Validity was assessed by comparing the clinical rating of psychosis with the CARS at the same evaluation session. The correlation obtained between the scale scores and clinicians' ratings was .84. Also, total scores were correlated with independent clinical assessments made by a child psychiatrist and child psychologist which resulted in a correlation of .80 (Schopler et al., 1980).

Discussion/Concerns: The CARS appears to be a good attempt at formalizing the process of diagnosing autism in children although it is an inherently subjective instrument. The scale appears to have adequate validity and reliability. However, interrater reliability, which is considered to be a very crucial component of a diagnostic tool such as this, is inadequate. Standardization procedures are also inadequate; however, due to the fact that finding an accessible population of children exhibiting autistic traits is difficult, standardization of the CARS has merit. A strength of this rating scale is its focus on direct observation with immediate rating. The authors caution that information derived from the CARS should not replace information from the child's history, home, school, and other experiences, but should be used as a supplement to other assessments.

References:

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, D. (1980). Toward objective classification of childhood autism: Childhood autism rating scale (CARS). *Journal of Autism and Developmental Disorders, 10*, 91-103.

Child Behavior Checklist (CBCL)

Author(s):	T. M. Achenbach
Publisher:	University Associates in Psychiatry
Address:	1 South Prospect St., Burlington, VT 05401 (802) 656-8313
Copyright Date:	1991/1992
Price:	CBCL/4-18 - approximately $48 (manual & checklists) CBCL/2-3 - approximately $40 (manual & checklists) (available 9/92)

Purpose: The purpose of the CBCL is to assess the competencies and problems of children and adolescents through the use of ratings and reports by different informants (Achenbach, 1991).

Description: The CBCL for ages 4 to 18 consists of both parent report forms and teacher report forms. (For adolescents ages 11-18, a youth report form is available.) The CBCL for ages two to three includes only parent report forms. For the CBCL/4-18 scales are included: (1) behavior problems scale (internalizing, externalizing, depression, etc.), and (2) social competence (social, school, and activities). The CBCL/2-3 includes six scales: aggressive behavior, anxious/depressed, destructive behavior, sleep problems, somatic complaints, withdrawn, and internalizing/externalizing.

Range of Children: CBCL can be used to evaluate children and adolescents from age 4 to 18. Parent report forms can be used for children ages 2-3.

Testing Time: It takes approximately 15 minutes to fill out the CBCL forms.

Scoring: After the forms have been completed, the CBCL can be either hand scored or computer scored. It uses a three-point rating scale from zero to two. Raw scores can be converted to T-scores and percentile ranks.

Examiner: Professional, usually a special education teacher or psychologist.

Standardization: Normative data for the CBCL/4-18 was based on a sample of 2,368 nonhandicapped children. The sample was representative with respect to age, sex, region, ethnicity, socioeconomic background, and community size. Data were obtained through home surveys involving both mothers and fathers. The normative sample was based on those children who had not received mental health services or special remedial education classes within the last year. Standardization information is not yet available for the CBCL/2-3.

Reliability: Test-retest reliability was assessed over one week. Reliability coefficients for the social competence scale and the social-emotional problems scale were .87 and .89 respectively. Interrater reliability between parents (mothers and fathers) ranged from .65 to .78 for both scales. "Intra-class correlations in the .90s for the mean item scores were obtained," according to Achenbach (1991). Reliability information for the CBCL/2-3 is not yet available.

Validity: Content validity is supported by the fact that items were selected for the CBCL because of significant relationships between the items and referrals for social-emotional problems. Construct validity is supported by the ability of the CBCL to discriminate between referred and non-referred children. Concurrent validity was assessed by comparing the CBCL to a number of rating scales. It correlated with the Connor's Parent Questionnaire at .82 with the Quay-Peterson Revised Behavior Checklist at .81. Criterion-related validity is also discussed in the manual and is reported as adequate by the author. Validity information for the CBCL/2-3 is not yet available.

Discussion/Concerns: A major strength of the CBCL is that it is fairly comprehensive and gathers information from multiple sources at least for those children who are in school. The CBCL also provides a means of assessing a child over time as it can be used from ages 2-18. The manual is very detailed and informative. The standardization of the CBCL appears to be adequate. The author provides a thorough discussion of validity in the manual but suggests that users of the CBCL should ultimately judge whether the content is appropriate for their purposes. It is also important to note that the scales were selected to summarize item content of each scale and are not to be used as diagnostic categories. Interrater reliability of this instrument is considered inadequate; therefore, caution should be exercised when using the CBCL for diagnostic purposes. The CBCL attempts to achieve an assessment-to-intervention link and is considered to be one of the better instruments in its category. It is especially useful when used in conjunction with other assessment tools.

References:

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist.* Burlington, VT: Thomas M. Achenbach.

Cognitive Abilities Scale (CAS)

Author(s):	Sharon Bradley-Johnson
Publisher:	Pro-ed
Address:	8700 Shoal Creek Boulevard, Austin, TX 78758-6897
Copyright Date:	1987
Price:	$109 (Complete Kit)

Purpose: The CAS was designed to provide detailed, educationally useful information on performance of young children in five areas relevant to later school success.

Description: The CAS consists of five subtests. The Language subtest contains 30 items and assesses a child's ability to understand and use oral language. The Reading subtest is comprised of 16 items and measures early reading skills such as naming letters. The Mathematics subtest (22 items) taps a variety of mathematical skills such as recognition of numbers and matching numbers. The Handwriting subtest (6 items) looks at skills directly involved in manuscript writing. The Enabling Behaviors subtest assesses abilities which are important for efficient learning, such as attending, imitating, following directions, etc. The CAS contains manipulable objects, pictures, and letters/numbers for the child to respond to during testing.

Range of Children: The CAS can be administered to children from ages 2-3 years old.

Testing Time: It is not a timed test; therefore, no time limits should be used. Children usually complete the CAS within 30 to 45 minutes.

Scoring: The child's responses are entered into a record book. Raw scores can be converted to standard scores, percentile ranks, and a cognitive quotient (M=100, SD=15). A nonvocal cognitive quotient and vocal cognitive quotient can also be obtained.

Examiner: Competent professionals with graduate training in individual assessment.

Standardization: The CAS was standardized on 536 children from 27 states. The sample was characteristic of the 1980 U.S. population relative to sex, residence, occupational status of parents, race, and geographic region. Mentally handicapped children were also included in the sample.

Reliability: Internal consistency coefficients for the subtests ranged from .80 to .94 at the two age levels. Total test coefficients were .97 for 2-year-olds and .96 for 3-year-olds. Test-retest reliability coefficients were highly variable ranging from .69 to .98 for subtests. Total test coefficients ranged from .90 to .99. Standard error of measurement is fairly low.

Validity: Concurrent validity was assessed by comparing the DAS to several tests. Correlation coefficients ranged from .63 to .84 when compared with five other tests. Predictive validity coefficients were obtained in two studies and are considered to be low. The author, in assessing construct validity, indicates that the CAS can differentiate between groups and that older children do obtain higher scores than younger children. Content validity is also discussed in the manual and is purportedly established.

Discussion/Concerns: The CAS appears to be an instrument which effectively assesses the skills of 2- and 3-year-olds as it focuses specifically on this age level and is fairly comprehensive. However, some of its psychometric properties are inadequate. Of particular concern is the limited standardization sample as well as low predictive validity coefficients. There is little independent research to support the author's contention about validity. The test itself appears highly attractive to young children as it contains colorful and interesting test items. Directions are well described and the manual is informative. The CAS appears to yield valuable information, but should be predominantly used as part of an assessment battery. Its ability to predict future academic functioning is somewhat questionable. There is little independent research to support the author's contention about the scale's usefulness.

References:

Bradley-Johnson, S. (1987). *Cognitive Abilities Scale*. Austin, TX: Pro-Ed.

Cognitive Skills Assessment Battery, Second Edition

Author(s):	Ann E. Boehm and Barbara R. Slater
Publisher:	Teachers College Press
Address:	Teachers College, Columbia University, New York, NY 10027
Copyright Date:	1981
Price:	$51.95

Purpose: The Cognitive Skills Assessment Battery was designed to provide a profile of strengths and weakness of prekindergarten and kindergarten children's cognitive and physical motor skills for the purpose of curriculum planning in prekindergarten and kindergarten programs.

Description: The second edition, according to the authors, responds to user comments. The test construction and scoring procedure are the same; however, some items were redrawn, color production was improved, and directions were clarified.

The authors claim that the battery is criterion referenced although criterion levels are not set, and total scores are not obtained (Diamond, 1989). The battery is individually administered and assesses skills in five major areas relevant to kindergarten and first grade curricula: (1) Orientation Toward's One's Environment, (2) Discrimination of Similarities and Differences, (3) Comprehension and Concept Formation, (4) Coordination, and (5) Immediate and Delayed Memory. The authors recommend that the test battery is best used for planning and guiding instruction.

Range of Children: Prekindergarten and kindergarten

Testing Time: 20-25 minutes

Scoring: The child's responses are coded as either plus or minus or by three different levels of competency. In addition, the child's behavior is coded during the task, on a scale of 1-4, on eight behaviors such as task persistence, attention span, and attention to directions.

Examiner: Teachers, aides, school psychologists, and learning disability specialists.

Standardization: The authors report that the sample selection was representative of the U.S. population with regard to sex, race, SES, community size, and native language, but this sample cannot be presumed to be representative because the authors did not report how the sample was selected or national data for comparing the characteristics. Eight hundred and sixty prekindergarten and kindergarten children were tested in the early fall of 1979 and 558 prekindergarten and kindergarten children in the late spring of 1980.

Reliability:

Information about the test's reliability is incomplete and must, therefore, be judged as inadequate. Although the test-retest reliability coefficient is adequate, with an average agreement of .80, a small sample (16 preschool and 32 kindergarten children) was tested and retested over a three-week period.

In addition, due to the fact that inexperienced individuals may administer the test, interrater reliability should have been reported (Embretson, 1989). Internal consistency, standard error of measurement, and standard deviation were not reported either.

Validity:

Validity is not adequately established by the author, only content validity is reported. Scale items were obtained by review of curricular materials, teacher interview, classroom observations, reviews of existing tests, review of the relevant research literature, and field testing.

Discussion/Concerns:

The manual provides many helpful suggestions and caveats for use of the test results. The caveats should be heeded by the test user. According to reviewers Embretson (1989) and Diamond (1989), due to the fact that the test lacks important technical data, there is not justification for its use as anything more than a rough, informal assessment tool.

References:

Diamond, E. E. (1989). Review of Cognitive Skills Assessment Battery. In J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (Vol. I, pp. 355-356). Lincoln: University of Nebraska Press.

Embretson, S. (1989). Review of Cognitive Skills Assessment Battery. In J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (Vol. I, pp. 356-357). Lincoln: University of Nebraska Press.

48

Developmental Profile II (DP-II)

Author(s):	Gerald Alpern, Thomas Boll, & Marsha Shearer
Publisher:	Western Psychological Services
Address:	12031 Wilshire Boulevard, Los Angeles, CA 90025
Copyright Date:	1986
Price:	$85

Purpose: The Developmental Profile II is an inventory of skills designed to assess a child's functional and developmental level.

Description: The Developmental Profile II was revised in 1986. This current revision of items was based upon feedback from users, but did not include a restandardization. The modifications included deleting items above the age of 9 years, 6 months; clarifying instructions; and removing sexist items and language. The revised Developmental Profile II consists of 186 "yes-no" items administered to parents or teachers as a direct test or by interview. It is designed to assess a child's functioning in the following five developmental areas: physical, self-help, social, academic, and communication.

Range of Children: The authors report that the Developmental Profile II assesses developmental behavior from birth to 9½ years, although the functional utility of the scale goes only to about age seven. Also, the authors state that the scale is appropriate for handicapped children of any age, although psychometric information is not included in the manual about this population.

Testing Time: 20-40 minutes

Scoring: Raw scores for each area are converted into age scores. Percentile ranks and standard scores are not available. The Academic Scale score can also be converted into a ratio IQ score; however, it should not be used because ratio IQ scores are obsolete and misleading and should not be based on third-party report data.

Examiner: It may be administered by teachers, teacher aides, physicians, nurses, social workers, medical aides, psychologists, etc.

Standardization: The standardization sample has several limitations with regard to generalizability. The normative sample is adequate with regard to size (3,008) but is limited with regard to distribution. Ninety-one percent of the sample was from the state of Indiana and nine percent from the state of Washington. The authors purposely excluded children with handicaps to represent normal developmental expectations. In addition, the standardization data are from the 1970 census and should be updated to reflect the current population. These limitations are mentioned by the authors of the test.

Reliability:

The reliability of the test is inadequate. The authors report test-retest reliability and internal consistency. The test-retest coefficients were low but adequate, ranging from .78 to .87; data were derived from an extremely small sample of 11 mothers, during an inadequate interval of only two to three days. A larger sample size and longer interval between testing are needed.

Based on a sample of over 1000 children, internal consistency coefficients ranging from .78 to .87 were obtained. These are low but adequate. These coefficients were determined on a modified version of the Profile.

Validity:

According to Hightower (1989), validity evidence is also inadequate. The authors provide various validity indices, but most of the studies provided have limited samples or are from earlier editions of the scale. The authors report, "there is limited support for convergent validity" and "the high correlations within scales may reflect a strong method bias."

Discussion/Concerns:

According to reviewer Hightower (1989), the Developmental Profile II is strong on clinical interpretations and guidelines, but has weak psychometric foundations to support the suggestions. The standardization sample is limited to two states, which adversely affects the generalization of results. The number of items within each age is small, generally two or three items. It seems reasonable that this instrument could be used to get a quick broad picture of a child's functioning, but that more psychometrically sound instruments would be needed for decision making or instructional planning (Huebner, 1989).

References:

Hightower, D. (1989). Review of Developmental Profile II. In J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (Vol. I, pp 249-250). Lincoln: University of Nebraska Press.

Huebner, E. S. (1989). Review of Developmental Profile II. In J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (Vol. I, pp 250-251). Lincoln: University of Nebraska Press.

Developmental Programming for Infants and Young Children (5 volumes)

Author(s):	D. Sue Schafer, Martha S. Moersch, & Diane B. D'Eugenio
Publisher:	The University of Michigan Press
Address:	PO Box 1104, Ann Arbor, MI 48106
Phone:	(313)764-4392
Copyright Date:	1981 (Spanish Vol II, 1988)
Price:	$43.45 set (Volumes may be purchased separately)

Purpose:

The five-volume set allows the educator/therapist to develop comprehensive and individualized developmental programs by translating comprehensive evaluation data rendered by the profile into short-term behavioral objectives which form the basis of daily activities planned to facilitate emerging skills.

Description:

Volumes I-II have been modestly revised since they were first published in 1977 in accordance with user comments. A Spanish version of Volume II has been added, although it must be ordered separately. Volumes I and IV comprise the assessment and application portions of the instrument. Volumes II and V consist of score sheets and developmental profiles. Volume III contains ideas for carrying out planned program objectives.

The instrument has six scales which provide developmental milestones in the following areas of development: perceptual/fine motor, cognition, language, social/emotional, self-care, and gross motor. The resulting profile indicates which skills are expected to emerge next in the child's development. There are no materials included; the examiner must put together an appropriate set, as described in the manuals. Volume IV does contain some pictures, and other paper materials will need to be prepared prior to administration.

Range of Children:

Volumes I-III are appropriate for children functioning in the 0-36-month range and Volumes IV-V are appropriate for children functioning in the 36- to 72-month range. The scales are appropriate for children who demonstrate a variety of handicapping conditions (mental retardation, hearing impaired, visually impaired, motorically involved).

Testing Time:

30-60 minutes

Scoring:

Items are scored in the following manner according to specified criteria: pass (P); pass-fail (PF); and fail (F). Items may be omitted (O). The child's performance on each of the six scales is then plotted on the Profile Graph.

Examiner: A professional. The authors recommend a multidisciplinary team which would include a psychologist or special educator, physical or occupational therapist, and a speech/language clinician.

Standardization: The profile has not been standardized on either a handicapped or a nonhandicapped sample of children. Assignment of items to specific age ranges was based on standardization or research from other instruments.

Reliability: The reliability seems to be adequate although small samples were used for the calculations. Interrater reliability was determined using a tester-observer method. The mean percent agreement between the tester and observers was 89 percent with a range from 80 percent to 97 percent.

Test-retest reliability was computed for 15 children who were administered the profile three times at three-month intervals. Correlations between the initial scores and the three-month retest ranged from .93 to .98. Correlations between the initial scores and the six-month retest ranged from .90 to .97.

Validity: Concurrent validity was examined by correlating the six profile scales with standardized instruments. The data are not particularly informative because the measures were administered to a small sample (14 handicapped children), the correlations had an extreme range (.33 to .96), and it is not known which correlations go with which set of measures. The following measures were administered: Bayley Mental and Motor Scales, Vineland Social Maturity Scale, Receptive-Expressive Emergent Language Scale (REEL), Clinical Motor Evaluation, and Slosson.

Discussion/Concerns: Developmental Programming for Infants and Young Children appears to be a useful combination of assessment items and programming activities for professionals working with young handicapped children. Many items included in the scales reflect current theories in the areas of language, cognition, and social-emotional development rather than simply compiling items taken from older standardized profiles. They attempt to look at the functional aspects of the child's development rather than discrete isolated skill development. The psychometric properties of the test are weak. Although reliability and validity correlations are adequate, they must be used cautiously as the samples were limited in size and range. The authors frequently stress the appropriate uses for their instrument and point out its weaknesses. While not useful as a placement tool, the profile is useful for program planning.

References:

Schafer, S. D. & Moersch M. S. (Eds.), (1981). *Developmental Programming for Infants and Young Children* (5 volumes). Ann Arbor: The University of Michigan Press.

Developmental Test of Visual-Motor Integration—3rd Revision (VMI)

Author(s): Keith E. Berry
Publisher: Modern Curriculum Press
Address: 13900 Prospect Rd., Cleveland, OH 44136
Copyright Date: 1989
Price: $17.33 - manual
$48.46 - 25 long forms
$34.88 - 25 short forms

Purpose: The VMI is a screening instrument which assesses visual-motor deficits in children and adults. It was renormed in 1982 and revised in 1989.

Description: The VMI consists of 24 geometric forms, arranged in developmental sequence, from simple to complex. The child is asked to copy the forms into a test booklet. The authors view visual motor behavior as a composite of other behaviors, including visual perception and motor coordination. Therefore, techniques for determining specific areas of difficulty are provided. Teaching techniques to parallel areas of assessment are also provided. A short form, composed of 15 drawings, is often used with ages 3-8.

Range of Children: The VMI can be administered to children ranging in age from 2-15, but it was designed primarily for preschool and the early primary grades. Users have administered it to learning disabled, educable mentally retarded, emotionally disturbed, and hearing impaired students.

Testing Time: The VMI can be administered in approximately 10 to 15 minutes.

Scoring: After the child has copied the forms, the examiner scores the test items based on criteria which are available in the manual. Raw scores can be converted to percentile ranks and standard scores based on a mean of 10 and a standard deviation of 3.

Examiner: Professional (preschool teachers, primary teachers, special education teachers, clinicians).

Standardization: Standardization was accomplished by using a group of 3,090 children. The 1982 manual reports proportions based on ethnicity, income level, sex, and residence type, but not on geographic location.

Reliability: Three kinds of reliability are reported: interscorer, test-retest, and internal consistency. Interscorer reliability coefficients ranged from .58 to .99 with a median of .93. A median internal consistency reliability of .79 was computed. Test-retest reliability ranges from .66 to .92 over five studies.

Validity: A number of validity studies are reported. One study indicates that the VMI correlates .50 with readiness, .89 with chronological age, and from .41 to .82 with scores on the Bender Visual Motor Gestalt Test.

Discussion/Concerns: Reviewers of the instrument feel that the behavior sampling is limited, although more items are included than are found on similar tests such as the Bender Visual Motor Gestalt Test or the Memory for Designs Test. The VMI has high reliability when compared to other measures of perceptual-motor skills, but information on validity seems to be inadequate. Prediction studies for various age levels would provide the kind of evidence that seems to be needed. Scoring procedures contain a fair degree of subjectivity. In general, it is an instrument that can be effectively used along with other assessment instruments.

References:

Salvia, J. & Ysseldyke, J. E. (1988). *Assessment in Special and Remedial Education* (Fourth Edition). Boston: Houghton Mifflin.

Diagnostic Inventory of Early Development—Revised (Brigance)

Author(s):	Albert H. Brigance
Publisher:	Curriculum Associates
Address:	5 Esquire Road, No. Billerica, MA 01862
Phone number:	1-800-225-0248
Copyright Date:	1991
Price:	$95

Purpose: The Brigance is designed to determine the developmental or performance level of the infant or child, to identify areas of strength and weakness, and to identify appropriate instructional objectives for the infant or child to develop an instructional program for infants and young children.

Description: The revised Brigance is the second edition of the Inventory which was originally published in 1978. The criterion-referenced, individually administered Inventory consists of over 200 developmental, readiness, and early academic skills in 11 major skills areas. The new edition includes a separate Social and Emotional Skills section. In addition, there are primary and secondary skills within each developmental skills section to be used for instructional planning for children with more severe developmental delays. During administration, the Inventory booklet is opened with the visual material facing the child. Materials to assess the child are provided by the examiner, and the authors recommend that they be materials which are familiar to the child and commonly found in the home or school.

Range of Children: Infants and children below the developmental level of 7 years.

Testing Time: Testing time will vary as the examiner selects the most appropriate skills and levels to administer to the individual child.

Scoring: A recordkeeping system, rather than a scoring system, is provided. The examiner circles each skill the child appropriately performs. A color-coding system is used to identify skills mastered and objectives identified at each time of evaluation to maintain an ongoing specific and easily interpreted record.

Examiner: Professional or paraprofessional

Standardization: The standardization sample was somewhat limited. The author reports that the sample is representative of different gender and ethnic groups and includes subjects from large urban, suburban, and rural locations. However, subjects were drawn from six U.S. cities, mainly on the west and east coasts, and one city in Canada.

Reliability:

Two types of reliability, test-retest and internal consistency, were determined for all sections and subtests. Test-retest reliability for each subtest and each age was basically sufficient, although some subtests and age groups did not have adequate reliability. Reliability ranged from a low of .58 (Section C for the 25-36-month range) to .99 in a large number of age range and section samples. Generally, coefficients for age ranges and sections were in the .90s.

Internal consistency is generally sufficient with coefficients in the mid to upper .90s. They range from a low of .44 in section B at 61-72 months of age, to a high of .99 for a number of age ranges and sections.

Validity:

The author reports that content validity is adequate. It was calculated using Hambleton's model for establishing item-objective congruence. "A congruence score of .90 or better is required to maintain an item in a criterion-referenced instrument and all items that are contained in the IED-R have received such a score." (Enright, 1991).

The Brigance proposes to assess developmental skills. To establish concurrent validity, the researchers determined that, if these skills were sequenced correctly and the sequences were indeed developmental, then the mean scores should increase across age groups. This was found to be true for the majority of the age groups and sections.

Discussion/Concerns:

The test is basically psychometrically sound. There are some test sections at a particular age level which have inadequate test-retest reliability and internal consistency. Results for these age levels should be interpreted cautiously. Overall, the test appears to be a comprehensive compilation of skills that young children demonstrate on a developmental basis. The inventory is useful for instructional planning, but, because it is a criterion-referenced instrument, it should not be used to make identification and placement decisions.

References:

Enright, B. E. (1991). *Brigance Diagnostic Inventory of Early Development-Revised (IED-R)*. Charlotte: University of North Carolina Press.

Differential Abilities Scale (DAS)

Author(s):	Colin D. Elliott
Publisher:	Psychological Corporation
Address:	555 Academic Court, San Antonio, TX 78204
Copyright Date:	1979, 1983, 1990
Price:	$495 per kit

Purpose: The DAS is an individually administered cognitive abilities test for children, preschool through high school.

Description: The DAS consists of 17 cognitive and three achievement subtests created to assess the multidimensional nature of children's abilities. At the preschool level, the achievement subtests are not used. The preschool level subtests are divided into three areas: Verbal Ability, Nonverbal Ability, and Diagnostic Subtests. The verbal and nonverbal ability areas measure a child's performance based on six subtests including Verbal Comprehension, Naming Vocabulary, Picture Similarities, Pattern Construction, Copying, and Early Number Concepts. They are intended to be used for placement and classification purposes. The Diagnostic Subtests include Block Building, Matching Letter-Like Forms, Recall of Digits, Recall of Objects, and Recognition of Pictures. These subtests are intended to be used to assess a child's strengths and weaknesses. A General Conceptual Ability (GCA) score can be obtained using various combinations of subtests based on the child's age. They are described in the manual. In general, the preschool scale assesses a child's verbal, reasoning, perceptual, and memory abilities. The DAS requires the child to respond to a number of different types of tasks.

Range of Children: The DAS can be used to assess children from ages 2-6 to 17. The DAS preschool level can be used with children from ages 2-6 to 7-11.

Testing Time: It takes approximately 35 minutes to an hour to administer this test.

Scoring: Subtests have recommended starting and stopping points based on the child's age. Raw scores can be converted to percentile ranks, standard scores, T-scores, and General Conceptual Ability Scores (M = 100, SD = 15).

Examiner: Professional (psychologist).

Standardization: The DAS was normed on a sample population of 3,475 children closely matching the 1988 U.S. Census figures for race/ethnicity, sex, region, community size, socioeconomic status, and parent education. Exceptional children such as learning disabled, mentally retarded, hearing impaired, etc., were included. Within each six-month age level, 175 to 200 children were included (Elliott, Daniel, & Guiton, 1991). Statistical analysis of test-bias was also conducted.

Reliability: Composite score mean coefficients of .90 and. 94 were obtained for internal consistency at two different ages levels. Test-retest reliability for composite scores ranged from .79 to .94 at two age levels; the majority were in the .85 to .95 range. Both internal consistency and test-retest reliability coefficients for subtests were significantly lower than for area scores.

Validity: Validity of the DAS was well-discussed by the author in the manual. In a review of the DAS, Elliott, Daniel, and Guiton (1991) discussed factor analytic studies which guided the structuring of the DAS. They also indicated that the DAS has been compared to several other ability tests, and correlation coefficients are generally in the high .70s to high .80s. For example, the DAS and WPPSI correlated at .89. Construct validity was also assessed and is discussed in the manual.

Discussion/Concerns: The DAS appears to be an adequate test psychometrically. The author states that the DAS's "most useful feature might prove to be the facilities it offers for tailoring content to the individual child's ability level to obtain maximum accuracy, while still providing normative interpretation." The test itself contains colorful objects and pictures which are attractive to young children. The *Introductory and Technical Handbook* and the *Administration and Scoring Manual* are comprehensive and informative. Overall, the DAS appears to be an instrument which can effectively assess a child's ability level and is considered to be equal to other ability tests.

References:

Elliott, C. D. (1990). *Introductory and Technical Handbook*. San Antonio, TX: Psychological Corporation.

Elliott, C. D. (1990). *Administration and Scoring Manual*. San Antonio, TX: Psychological Corporation.

Elliott, C. D., Daniel, M. H., & Guiton, G. W. (1991). Preschool Cognitive Assessment with the Differential Ability Scales. In B. A. Bracken (Ed.), *The Psychoeducational Assessment of Preschool Children* (pp. 133-153). Needham Heights, MA: Allyn and Bacon.

The Early Learning Accomplishment Profile (E-LAP) for Developmentally Young Children

Author(s):	M. Elayne Glover, Jodi L. Preminger, and Anne R. Sanford
Publisher:	Kaplan Press
Address:	PO Box 609, Lewisville, NC 27023
Copyright Date:	1989
Price:	$6 per profile, $39.95 Early-LAP Activity Cards, and $265 for E-LAP Developmental Kit

Purpose: The E-LAP was developed to generate developmentally appropriate instructional objectives and task analysis programming for the young handicapped child who requires a break-down of the developmental sequence into smaller steps.

Description: The E-LAP is an individually administered criterion-referenced tool that assesses a young child's development in the areas of Gross Motor, Fine Motor, Cognitive, Language, Self-Help, and Social/Emotional skill areas. There are 414 items taken from previously developed instruments. Items are stated as behavioral objectives. The E-LAP was revised in 1989, although not substantially. Some of the skills were changed or moved, and the scoring profile was made somewhat easier to use, but the technical data was not updated.

Range of Children: Birth to 36 months

Testing Time: No information provided.

Scoring: Each skill is recorded in one of three ways: (+) positive demonstration of skill, (-) no demonstration of skill, and (+/-) emerging skills according to specified criteria. Each skill is recorded in one of three columns: the "Pre" column (initial assessment), the "Post" column (end of program) or the "achievement date" column for ongoing assessment. Appropriate starting and discontinuing points are discussed in the manual. Approximate developmental ages may be obtained for each of the six areas.

Standardization: No norm group is reported, although many of the items were taken from other norm-referenced instruments.

Reliability: Only interrater reliability was calculated using a small sample of 11 males and seven females. Correlations of between .93 to 1.0 were reported between raters for the gross motor, fine motor, language, and cognitive sections of the test.

Validity: Fourteen children were administered the Bayley Scales of Infant Development to validate the item selection of the E-LAP. Concurrent validity was adequate. A combination of the fine motor and gross motor skills on the E-LAP had a correlation of .85 with the Bayley Motor Scale, and a combination of the E-LAP language and cognitive skills had a correlation of .93 with the Bayley Mental Scales.

Discussion/Concerns: The E-LAP is not appropriate for decision-making purposes. Reliability and validity data are limited due to the small sample size and the number of studies conducted. Due to the lack of a standardization sample, the developmental ages should be reported *only* when a rough estimate is required. The E-LAP seems to fulfill its intended purpose of providing developmental sequences which are broken into small steps for infants and severely/multihandicapped children. The E-LAP makes writing individual education plans easier as the steps between the skills are relatively small.

References:

Glover, M., Preminger, J. L., & Sanford, A., (1989). *Learning Accomplishment Profile for Developmentally Young Children (E-LAP).* Winston-Salem, NC: Kaplan Press.

60

Expressive One-Word Picture Vocabulary Test—Revised (EOWPVT-R)

Author(s):	Morrison F. Gardner
Publisher:	Academic Therapy Publications
Address:	20 Commercial Boulevard, Novato, CA 94949
Copyright Date:	1990
Price:	$75 per kit

Purpose: The EOWPVT-R assesses expressive language in children. It aids in estimating what a child has learned from his/her environment and from formal education. It can be used for readiness screening for preschool and kindergarten children as well as estimating bilingual students' fluency in English.

Description: The EOWPVT-R has been re-standardized, re-normed, and some items have been changed while the basic purpose and concept of this instrument has stayed the same. The EOWPVT-R is composed of 100 picture items. The pictures are shown to the child, who is asked to name things in the pictures, usually with one word. Basal and ceiling levels are included and the child's answers are recorded on an individual record form. Picture items are purported to be bias-free. Directions for administration to Spanish-speaking children are provided in the manual.

Range of Children: The EOWPVT-R can be administered to children from ages 2-0 to 11-11.

Testing Time: Administration takes approximately 15 minutes. Younger children may take less time.

Scoring: Raw scores can be converted to age equivalents, standard scores, scaled scores, percentile ranks, and stanines.

Examiner: Professional such as a psychologist or speech and language clinician.

Standardization: This instrument was standardized on a norm group of 1,118 males and females residing in the San Francisco Bay Area. It was normed on children who were predominantly English speaking. Children were tested in a number of school settings as well as in the home.

Reliability: Internal consistency reliabilities for this test ranged from .84 to .92 with a median reliability of .90. No other reliability information is provided in the manual.

Validity:

The manual discusses content, item, and criterion-related validity. The author included items which were representative of a common core of English words that are typically used by children. Also, items that were included yielded moderate to high positive discrimination and item-total correlations, both over the test as a whole and within each age group. Criterion-related validity was assessed by comparing the EOWPVT-R with the Peabody Picture Vocabulary Test-R (PPVT-R), WPPSI-R, and WISC-R. The EOWPVT-R correlates with the PPVT-R at .59. Correlations with the WPPSI-R and WISC-R were lower.

Discussion/Concerns:

Psychometric characteristics of the EOWPVT-R are considered inadequate mainly due to the limited standardization sample used for norming, the lack of information provided concerning test-retest reliability, and inadequate criterion-related validity. The manual indicates that the EOWPVT-R can be used to obtain a valid estimate of a child's "verbal intelligence"; however, scores derived from this test should not be interpreted in this manner. In general, this instrument could be effective when used as a screening instrument and/or as part of an assessment battery.

References:

Gardner, Morrison F. (1990). *Expressive One-Word Picture Vocabulary Test-Revised*. Novato, CA: Academic Therapy Publication.

62

The Fisher-Logemann Test of Articulation Competence

Author(s):	Hilda A. Fisher and Jerilyn A. Logeman
Publisher:	The Riverside Publishing Company
Address:	80420 Bryn Mawr Avenue, Chicago, IL 60631
Copyright Date:	1971
Price:	$75

Purpose: The Fisher-Logemann test of Articulation Competence is designed (1) to examine the test subject's phonological system in an orderly framework; (2) to provide ease in recording and analyzing phonetic notations of articulation; and (3) to facilitate accurate and complete analysis and categorization of articulatory errors.

Description: The picture version consists of 109 line drawings. Eleven of the cards are tab-indexed for screening test applications. Each picture is designed to elicit a spontaneous single-word response. The test assesses phonemes consisting of 25 consonants and 21 consonant blends.

Range of Children: Picture Test—Preschool to adult
Sentence Test—Grade 3 to adult

Scoring: The scoring procedure is designed to record not only instances in which the subject produces the phonemes of English acceptably by adult standards but also an analysis of the nature of the misarticulations. Consonant phonemes are analyzed on the basis of a three-feature system: place of articulation, manner of articulation, and voicing. Vowel phonemes are analyzed on the basis of a four-feature system: place of articulation, height and tongue, tension, and lip rounding.

Examiner: Speech/Language Clinician

Standardization: The Fisher-Logemann was standardized on approximately 500 children in the Chicago area. The children were representative of varying dialectal backgrounds, and the sample included children with various mental handicaps and speech problems. Subjects were of various ages; however, they were predominantly from the lower elementary grades.

Reliability: No reliability studies are reported in the manual.

Validity: No validity information is provided in the manual.

Discussion/Concerns: It is impossible to evaluate the Fisher-Logemann in terms of technical adequacy; however, it can be considered a useful instrument for limited purposes. The instrument provides a sample of an individual's sound production and yields information for program planning. The manual is informative; however, scoring of the Fisher-Logemann appears somewhat cumbersome. Overall, this instrument should not be used as a diagnostic instrument and should be used with other instruments when assessing a child's articulation competence.

References:

Fisher, H. A. & Logemann, H. A. (1971). *The Fisher-Logemann Test of Articulation Competence*. Geneva, IL: Riverside Publishing.

64

Revised Gesell and Amatruda Developmental and Neurologic Examination

Author(s): Hilda Knobloch, Frances Stevens, and Anthony F. Malone
Publisher: Developmental Evaluation Materials, Inc.
Address: PO Box 272391, Houston, TX 77277-2391.
Copyright Date: 1987
Price: Contact publisher

Purpose: The Developmental Schedules were developed and organized to provide developmental diagnosis across several developmental areas.

Description: Knobloch, Stevens, and Malone have reorganized the 1940 (and later) compilations of Gesell's developmental schedules to reflect more accurately rates of growth of today's infants and young children. The revised Schedules examine the quality and integration of five different fields of behavior: (1) adaptive behavior, (2) gross motor behavior, (3) fine motor behavior, (4) language behavior, and (5) personal-social behavior. There are eight key age zones each with its own examination sequence. The Knobloch et al. manual devotes two chapters to the conduct of the examination and specifics of the examination procedures. Materials for the examination can be readily improvised or secured from other test kits. They may also be ordered from the publisher. In the revised schedules some sequences have been altered due to the accelerated acquisition of these infant behaviors, and items have been added and changed in various fields and age levels, especially in the 24- to 36-month age levels.

Range of Children: Four weeks to 36 months.

Testing Time: Variable; depending on the child, it may take under 10 minutes or up to 30 minutes or more.

Scoring: On the basis of history and observations, the examiner selects the appropriate schedule for the child. A coding key is used for scoring. A (+) is entered in the observation column when the behavior pattern is present, (+/-) is entered for emerging patterns, and (-) is entered when a pattern is absent. Various other codes which are included in the manual are used. The child's maturity level in each field of behavior is determined where the aggregate of + signs changes to an aggregate of - signs. These maturity age levels are the basis for calculating the developmental quotients (DQ) in each area. The following ratio is used to do so:

$$DQ = \frac{\text{Maturity Age}}{\text{Chronological Age}} \times 100$$

Examiner: A professional who is familiar with the schedules.

BEST COPY AVAILABLE

Standardization: The standardization sample consisted of "normal" (according to specified criteria) infants and children from the Albany, NY area in 1975 and 1977. Nine hundred and twenty-seven examinations were completed on 233 children and infants at 20 age levels from four weeks through 36 months. The sample is limited geographically and is not representative of the population of the United States.

Reliability: Inter-observer reliability was adequate. It was evaluated in two separate studies. The percentage of agreement between observers for individual items was determined on a sample of 48 cases, covering age ranges from 16 weeks to 21 months. The overall percentage of agreement for 2,302 comparisons of 305 behavior patters was 93 percent. In the second procedure, the reliability between two observers on 184 cases at nine age levels ranged from .84 to .99 with the majority of correlations above .90.

Validity: No information is reported

Discussion/Concerns: The Gesell Developmental Schedules are contained in the *Manual of Developmental Diagnosis*. The Schedules are the foundation for the majority of current developmental assessment instruments. They evaluated a young child's developmental status. Behaviors are adequately described in the manual and easy to understand. According to the author, the Schedules are an adequate diagnostic measure when used by a physician. However, when used by other professionals they should only be used to evaluate the child's developmental status. The manual is informative; behavior sequences are pictured and described in detail. The reader is instructed in the construction of needed materials or told how they may be obtained. It is suggested that the examiner be thoroughly familiar with the procedures before administration. The Schedules have been renormed, although the sample was limited to one city and its surrounding communities, and 15 years have since elapsed. The Schedules will be more useful for program planning than for placement decisions.

References: Knobloch, H., Stevens, F., & Malone, A. F. (1987). *Manual of Developmental Diagnosis*. Houston, TX: Developmental Evaluation Materials.

Goldman-Fristoe Test of Articulation (GFTA)

Author(s):	Ronald Goldman and Macalyne Fristoe
Publisher:	American Guidance Service
Address:	Circle Pines, MN 55014
Copyright Date:	1986
Price:	$94.50

Purpose: To sample a wide range of an individual's articulatory skills and to provide a systematic means of assessing an individual's articulation of the consonant sounds.

Description: The GFTA-R is an individually administered, criterion-referenced instrument. It has been revised, but construction of the test has changed little from the 1972 version with the exception of the addition of normative data for the 2- to 6-year-old age group.

The test is comprised of three subtests: Sounds in Words, Sounds in Sentences, and the Stimulability subtest. The Sounds in Words subtest is administered by asking the child to name a picture; the examiner records the subject's production of specific speech sounds in the initial, medial, and final positions in words. The Sounds-in-Sentences subtest consists of two narrative stories accompanied by action pictures. After reading the stories, the examiner asks the subject to retell the stories; information regarding the subject's articulatory skills is recorded. The stimulability subtest assesses the ability of the subject to correctly produce a previously misarticulated phoneme when given maximum stimulation, both visual and oral.

Range of Children: Two years and up.

Testing Time: 10-15 minutes for the Sounds-in-Words subtest. The administration times for the Sounds-in-Sentences and Stimulability subtests are not reported.

Scoring: Responses are scored during the test administration. This provides a profile of the subject's articulatory performance. Several types of articulation errors and interpretations are available in the manual.

Examiner: Speech/language clinician

Standardization: Normative data are provided for two groups: 2- to 6-year-olds and individuals 6 and older. These studies were not performed by the authors but were taken from other researchers' work using their instrument.

The Sounds-in-Words subtest was administered to 852 male and female children 2- to 6-years old from 41 sites in seven U.S. cities. Subjects were placed in one of four same-aged groups that were balanced to reflect the U.S. population according to sex, geography, and ethnic background. Children with obvious physical abnormalities were eliminated from the sample.

The Sounds-in-Words subtest was administered to 802 subjects in the six and older age group in a 1969 National Speech and Hearing Survey. It is not clear how norms were derived for the Stimulability subtests.

Reliability: Reliability of the GFTA is low. Test-retest reliabilities were established for the Sounds-in-Words Subtest and the Sounds-in-Sentences Subtest from data gathered by eight "experienced speech pathologists" who tested and retested 37 children between 4-8 years of age. Test-retest reliability for the Sounds-in-Words Subtest was quite high with a median agreement of 95 percent; for the Sounds-in-Sentences Subtest, it was 70 percent which is quite low.

Interrater reliability was 92 percent, but it was based on an extremely limited sample.

Validity: No information is available.

Discussion/Concerns: The GFTA is relatively easy to administer and score. It provides a fairly complete picture of a child's articulation skills. The authors point out the limitations and strengths of the test within the manual. The items within given age ranges, especially the Infant scale, are generally few—two or three items. According to Mowrer (1989), although it has definite psychometric limitations, the GFTA compares favorably to other similar picture articulation tests that check sounds in three positions. This instrument is adequate as a way to obtain a broad picture of the child's functioning, but it is not recommended for instructional planning.

References:

Mowrer, D. E. (1989). Review of Goldman Fristoe Test of Articulation. J. Close Conley & J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. (Vol. I, pp. 323-325). Lincoln: University of Nebraska Press.

Hawaii Early Learning Profile (HELP)

Author(s):	Setsu Furuno, Katherine A. O'Reilly, Carol M. Hosaka, Takayo T. Inatsuka, Toney L. Allman, & Barbara Zeisloft-Falbey
Publisher:	Vort Corporation
Address:	PO BOX 60132, Palo Alto, CA
Copyright Date:	1985
Price:	$2.45/set of 3 (HELP Charts), $19.95 (Activity Guide), $2.45/set (Checklist)

Purpose:

The HELP was developed to provide a comprehensive picture of the level of functioning of children with a wide range of handicaps and provides suggestions for activities to teach developmental skills, building on strengths as well as weaknesses.

Description:

The HELP consists of the HELP Activity Guide, HELP Checklist, and the HELP Charts. The HELP is not standardized. It was revised from the 1978 version based on field-testing and reviewer information to provide a month-to-month sequence of normal developmental skills in six different areas. The Charts display 650 developmentally sequenced skills to help identify current mastery of skills to identify target objectives, and to record and visually track progress in six areas: (1) Cognitive (2) Gross Motor (3) Fine Motor (4) Language (5) Social, and (6) Self-help.

The Checklist covers the same six developmental areas, but the format is different. The Checklist groups skills by areas and in age sequence with columns for easy recording of assessment dates, progress information, and comments on home involvement/support and case management. The checklist then aids in selecting individual educational objectives.

The Activity Guide offers task-analyzed activities and intervention strategies that correspond to the 650 HELP skills listed on the Charts.

Range of Children:

Handicapped Infants, toddlers, and developmentally delayed young children ages birth through 36 months.

Testing Time:

No information is provided.

Scoring:

On the Checklist and Charts, skills are assessed according to specific criteria. Appropriate coding symbols are reported in the manuals. Skills that are assessed as emerging or not observed can be targeted for intervention.

Examiner:

No special qualifications.

Standardization:

The HELP is not standardized. The Charts and Activity Guide were field-tested by all programs for infants with handicaps in Hawaii. Additionally, the materials were used and reviewed by programs in 35 states and seven different countries.

Reliability: No information is provided.

Validity: No information is provided.

Discussion/Concerns: The HELP system provides a variety of assessment instruments and intervention strategies. It lacks appropriate technical information and, therefore, should only be used for its stated purposes and not for making placement decisions.

References:

Furuno, S., O'Reilly, K. A., Hosaka, C. M., Inatsuka, T. T., Allman, T. L., & Zeisloft, B. (1985). *HELP*. Palo Alto, CA: Vort Corporation.

Kaufman Assessment Battery for Children (K-ABC)

Author(s): Kaufman, Alan S. and Kaufman, Nadeen L.
Publisher: American Guidance Service
Address: Publishers' Building, Circle Pines, MN 55014
Copyright Date: 1983
Price: $269—Complete Battery

Purpose:

The Kaufman Assessment Battery for Children is an individually administered measure of intelligence and achievement. "The K-ABC is intended for psychological and clinical assessment, psychoeducational evaluation of learning disabled and other exceptional children, educational planning and placement, minority group assessment, preschool assessment, neuropsychological assessment and research" (Administration and Interpretive Manual, 1983).

Description:

The K-ABC was developed from recent research in cognitive psychology and neuropsychology. It claims to assess simultaneous and sequential mental processing as well as school achievement (which measures a child's acquired knowledge). The authors indicate that the K-ABC is useful for the assessment of preschool children, elementary age school children, minorities, and exceptional children.

The K-ABC yields scores for four global areas: Sequential Processing, Simultaneous Processing, Mental Processing Composite (Simultaneous plus the Sequential), and Achievement. There are 16 subtests: 3 subtests in the Sequential Processing Scale; 7 subtests in the Simultaneous Processing Scale and 6 subtests in the Achievement Scale. A maximum of 13 subtests are administered to any one child. For 2½-year-olds, 7 subtests are used; 5-year-olds are given 12 subtests.

The K-ABC has supplementary sociocultural norms for minority children and has a special Nonverbal Scale which can be administered in pantomime to nonverbal children.

A unique feature of the K-ABC is that the examiner is allowed to teach the first three items of each mental processing subtest. Since the K-ABC was standardized with this unique feature, it is therefore acceptable.

The K-ABC is relatively easy to administer and provides objective scoring procedures. The authors indicate that the K-ABC yields scores which can be translated into teaching objectives and educational intervention strategies and is sensitive to the unique needs of preschool, minority, and exceptional children.

Range of Children: Children from 2½ to 12½ years.

Testing Time: The authors state that the test takes about 35 minutes at age 2½, 50 to 60 minutes at age 5, and 75 to 85 minutes at ages 7 and above.

Scoring: The K-ABC yields standard scores ($\underline{M}$ = 100 and $\underline{SD}$ = 15) for all four major scales. National percentile ranks, sociocultural percentile ranks, and age and grade equivalent scores can be computed also. The mental processing scaled scores have a mean of 10 and a standard deviation of 3. The Achievement subtests have a mean of 100 and a standard deviation of 15.

Tables in the Administration and Scoring Manual are easy to use and scoring is completed rapidly.

Examiner: The authors indicate that administration of the K-ABC requires an examiner who is well trained in psychology and individual intellectual assessment, and who has studied carefully both the K-ABC Interpretive Manual and K-ABC Administration and Scoring Manual.

Standardization: The national standardization sample consisted of more than 2,000 children tested in 34 test sites in 24 states. The norm group was stratified using the 1980 census data. The sample group was stratified within each age group by sex, geographic region, socioeconomic status, race or ethnic group, community size, and educational placement of the child (normal or special class). Exceptional children were systematically included In proportions representative of the population at large.

Reliability: Split-half reliability coefficients showed good internal consistency as mean values of .80 and above were obtained for 12 of the 16 subtests on the K-ABC. Internal consistency coefficients for the K-ABC Global Scales range from .86 (Simultaneous) to .93 (Achievement) for preschool children. The scale is somewhat weaker at the youngest ages (2½ to 3) but still well within acceptable limits. Test-retest reliability coefficients on the Mental Processing Scales range form .77 to .97 and improve with increasing age.

Standard Errors of Measurement for the K-ABC Mental Processing subtests range from approximately 1 to 1.5 points and for the Achievement subtests form approximately 4 to 6 points. Overall, the K-ABC is a statistically reliable instrument.

Validity: The authors report good construct validity for the K-ABC. Mean scores on the K-ABC increase steadily as age increases. Factor analytic studies support the two distinct types of mental processing scales (Sequential and Simultaneous) identified by the K-ABC.

More than 40 correlational studies reported in the administrative manual indicate satisfactory construct, concurrent and predictive validity. These studies compared the K-ABC with such well-known intelligence tests as the WISC-R and Stanford-Binet as well as other currently popular and accepted tests.

Discussion/Concerns:

The authors indicate that the K-ABC is predicated on a new definition of intelligence that defines intelligence as how children process information and solve problems as opposed to simply assessing the acquired knowledge of children. Therefore, the K-ABC is process-oriented rather than content-oriented, as are most other conventional tests of intelligence. The Mental Processing Scales are purported to be related to the fluid abilities of the Cattell-Horn theory, while the Achievement Scale is more closely related to crystallized abilities; however, a number of reviewers question this contention.

Statistical properties of the K-ABC, reported in the manual, are very good. While the K-ABC may be an appropriate choice for school-aged children, a number of reviewers question its appropriateness for preschoolers. Sattler (1988) points out that since only five subtests are used to form the Mental Processing Composite for 2½- to 3-year-olds, insufficient data for decision-making purposes are available. Bracken (1987) suggests that the K-ABC has an inadequate floor and steep item gradients and therefore doesn't discriminate handicapped from non-handicapped 2-, 3-, and 4-year-olds. Danielson (1989) concludes, in her review, that the K-ABC is a poor diagnostic choice since it fails more of the criteria of technical adequacy than do the other instruments available in this class of instruments.

References:

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment*, *4*, 313-326.

Danielson, E.B. (1989). Assessment of Young Children. *Childhood Issues*. A. Canter (Ed.), (pp. 11-18). Minneapolis: Minnesota Association of School Psychologists.

Kaufman, A. S. and Kaufman N. L. (1983). *Kaufman Assessment Battery for Children—Interpretive Manual and Administration and Scoring Guide*. Circle Pines, Minnesota: American Guidance Service.

Mitchell, J. V. Jr. (1985). *Ninth Mental Measurements Yearbook*. Lincoln: University of Nebraska Press.

Sattler, J. M. (1988). *Assessment of Children* (3rd Edition). San Diego: J. M. Sattler.

Learning Accomplishment Profile, - D Standardized Assessment (LAP-D)

Author(s):	David Wilson LeMay, Patricia M. Griffin, and Anne R. Sanford
Publisher:	Kaplan School Supply Corporation
Address:	Box 609, Lewisville, NC 27023
Copyright Date:	1992
Price:	$450 per kit

Purpose: To evaluate the entry program skills of a child in order to develop an appropriate intervention and to evaluate the exit skills of a child in order to validate the effects of the intervention program.

Description: The LAP-D has been revised from the previous edition. The Self-Help Scale and its five subscales have been discontinued, some skills and the corresponding developmental ages have been changed, and developmental ages have been raised or lowered in some cases. Items were taken from previously developed instruments. Four developmental skill areas and eight subscales are included: Fine Motor (Manipulation and Writing), Cognitive (Matching and Counting), Language (Naming and Comprehension), and Gross Motor (Body Movement and Object Movement). Each subtest is contained in a separate easel and the kit includes eight copies of the manual to facilitate "station" administration.

The scale is now standarized and norm-referenced and individually administered. Items are arranged in order of complexity and in task-analysis manner. Each item describes the behavior observed, the procedure to be followed in eliciting the desired response, and the criteria against which success is measured. Developmental age equivalents, percentile ranks, and Z scores are provided. The LAP-D kit contains administration directions and lists the materials needed to administer the assessment.

Range of Children: Children within a range of 30 to 60 months. The authors discuss its usefulness with children with a variety of developmental difficulties.

Testing Time: The average administration time reported by users is 1 to 1½ hours.

Scoring: Items are numbered in the Scoring Booklet and correspond to the items in the loose-leaf easel. If the child passes an item, a plus (+) is recorded, and if the item is failed, a minus (-) is recorded. A basal level is determined at the point at which the child has received three consecutive passes. The ceiling is the point at which the child fails three out of five items.

Examiner: Professional (teacher)

Standardization: Standardization procedures are available in the manual.

Reliability: Reliability data are available in the manual.

Validity: Validity data are available in the manual.

Discussion/Concerns:

References:

Lemay, D. W., Griffin, P. M., & Sanford, A. R. (1992). *Learning Accomplishment Profile (LAP-D)*. Lewisville, NC: Kaplan School Supply.

Leiter International Performance Scale

Author(s): Russell Graydon Leiter
Publisher: Stoelting Company
Address: 620 Wheat Lane, Wood Dale, IL 60191
Copyright Date: 1948
Price: $610

Purpose: The Leiter was designed as a non-verbal, individually administered intelligence test.

Description: The complete Leiter consists of 54 non-verbal subtests contained in three separate trays. The author indicated that it is suitable for a wide variety of subjects—ranging from the mentally retarded to the mentally superior. The Leiter does not require the examiner to use language to administer the test. The test does not require a verbal response from the examinee. The Leiter requires minimal recording which affords the examiner more opportunity to observe examiner behavior. Timed tests are negligible, thus making the Leiter a power, rather than a speed test.

Testing materials for the Leiter consist of a number of response blocks, stimulus strips, and an adjustable wooden frame. The examiner places the appropriate stimulus strip on the frame and the child inserts the proper blocks into the stalls of the frame. Difficulty increases at each age level. The Leiter is an age scale with four tests at each year level from year II through year XIV and six tests at year levels XVI and XVIII.

Range of Children: Ages 2 through Adulthood.

Testing Time: One to one and a half hours.

Scoring: The examiner scores the child responses during the assessment on a score card. A basal test age and terminal test age are determined for each child assessed. The Leiter yields intelligence quotients and mental ages. The Handbook by Levine (1982) has tables that contain all the adjustments recommended by Leiter. This section of the manual should be carefully reviewed by examiners. An adaptation of the Leiter was completed by Arthur in 1952 for use with young children ages 3-0 to 7-11.

Examiner: Psychologist-professional.

Standardization: Leiter norms were based on a Hawaiian sample in the 1940s and the Arthur Adaptation on an Illinois sample. This is reported as a major weakness of the Leiter. Examiners should use the Leiter carefully in light of this problem.

Reliability: The reliability of the 1948 version indicates that internal consistency ranges from .75 to .94 by various researchers. Test-retest reliability ranges from .36 to .92. The reliability data were accumulated from 1952 through 1973 and are reported in Levine's Handbook.

Validity: Twelve different studies have compared the Leiter and the Stanford-Binet. Correlations between the Leiter and the Stanford-Binet range from .38 to .89. The median correlates for both IQ and mental age is .77. For all forms of the Wechsler scales, the Leiter or Arthur adaptation IQ had median correlations of .78 with Full Scale IQs, .68 with Performance IQs, and .71 with Verbal IQs. A complete list of research on the Leiter as compared to other assessment instruments can be found in Levine's Handbook.

Levine concludes that the Leiter and the Arthur Adaptation of the scale have been shown to be positively and significantly correlated with a number of other criteria. The Leiter and Arthur Adaptation and Stanford-Binet appear to be moderately and positively correlated while they seem highly and positively correlated with the Wechsler Scales. Moderate correlations between the Leiter and Arthur Adaptation and teacher ratings of intelligence have been found.

Discussion/Concerns: The Leiter International Performance Scale is an individually administered nonverbal test of intelligence. The norms are very old. While a number of researchers have talked of renorming the Leiter, no such project has been completed.

Validation studies comparing the Leiter and Arthur Adaptation with other established instruments have been conducted and tend to show moderate to high correlations with other popular test instruments. Studies of the use of the Leiter with special populations, e.g., Hispanics, Blacks, mentally retarded, brain-damaged and physically handicapped are reviewed by Levine. The Leiter shows promise in its potential clinical usefulness with special populations such as speech/language and physically impaired individuals.

At this time, it would appear that the Leiter should be used as a part of the complete test battery rather than a single cognitive measure. It is useful when other alternatives do not exist, especially for 2- and 3-year-olds who do not have language skills.

References:

Leiter, R. G. (1980). *Leiter International Performance Scale—Instruction Manual*. Chicago: Stoelting.

Levine, M. N. (1982). *Leiter International Performance Scale: A Handbook*. Los Angeles: Western Psychological Services.

McCarthy Scale of Children's Abilities

Author(s):	Dorothea McCarthy
Publisher:	The Psychological Corporation
Address:	555 Academic Court, San Antonio, TX 78204
Copyright Date:	1972
Price:	$412.50 for Complete Set

Purpose:

The McCarthy Scales of Children's Abilities were designed to evaluate children's general intellectual level as well as their strengths and weaknesses in a number of ability areas.

Description:

This individually administered test consists of 18 subtests that make up six scales: Verbal, Perceptual-Performance, Quantitative, Memory, Motor, General Cognitive. The General Cognitive Scale is a composite of the Verbal, Perceptual-Performance and Quantitative Scales.

The test materials are attractive and help to facilitate the establishment and maintenance of rapport. The test begins with manipulative items; gross motor tests appear midway when the child is becoming restless and interest is beginning to wane; and the final items require limited vocalization in anticipation of the child's state of fatigue.

The McCarthy includes several built-in precautions to promote optimum measurement of the child's ability in each task. Extra trials are permitted for many items to give the child a second chance; only the best performance is counted. To ensure that the child understands the task at hand, he or she is frequently given feedback on the easier items; on one subtest, examples are given to get the child started. In addition, the inclusion of several multipoint items rewards the child for virtually any response approximating correct performance.

Range of Children:

The test can be administered to children ranging in age from 2½ to 8½.

Testing Time:

Approximately 45 minutes for children under 5 years and approximately 1 hour for older children.

Scoring:

Four kinds of scores are obtained: a General Cognitive Index, scale indexes, percentile ranks, and mental age. The author states that the General Cognitive Index "is a scaled score; it is not a quotient." The score has a mean of 100 and a standard deviation of 16. Separate indexes are obtained for each of the other five major scales; they have a mean of 50 and a standard deviation of 10. Tables in the manual are used to transform scaled scores into percentile ranks and to provide estimated mental ages for performance on the general cognitive scale.

Examiner:

Professional (psychologist)

Standardization: The McCarthy was standardized on a representative sample of 1,032 children. The sample was stratified on the basis of sex, age, race, geographic region, father's occupation, and urban-rural residence. Proportions in the normative sample approximate very closely the 1970 U.S. census data.

Reliability: Internal-consistency coefficients were completed for all but three subtests. Reliability coefficients for the six major scales show variation in adequacy. Reliabilities for the Verbal and General Cognitive scales meet acceptable standards. Internal consistency for the GCI averaged .93 across age groups.

Test-retest reliability coefficients obtained were .90 for the General Cognitive Index and an average of .81 for the five Scale Indexes. A stability coefficient of .85 was computed for the General Cognitive Index over a period of one year. However, sample sizes were small and included few children under age 5 (Paget, 1985).

Validity: To establish predictive validity, 31 children were tested using the McCarthy and then tested four months later using the Metropolitan Achievement Test (MAT). Correlations among the six scales of the McCarthy and the six scales of the MAT were high for the Perceptual-Performance and Quantitative scales, low for the General Cognitive Scale (.49) and poor for the Verbal, Memory and Motor Scales. The last three McCarthy scales had correlations averaging .15, .26 and .03 with the MAT. The author states that the results should be interpreted with caution because of the small size of the sample.

To establish concurrent validity, McCarthy scores were correlated with scores obtained on the Stanford-Binet and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). The sample consisted of 35 white children enrolled in a Catholic school in New York City. The McCarthy correlates .81 with Stanford-Binet and from .62 to .71 with the WPPSI. According to Paget (1985), construct validity evidence is adequate.

Discussion/Concerns: The McCarthy has become a popular test for assessing the abilities of preschool children. Its strengths are as follows: it has adequate psychometric properties; it yields reliable and stable global scores; there is empirical support for the interpretation of scores from the GCI, verbal, motor, and perceptual-performance scales; it correlates strongly with school achievement; it is fairly nondiscriminatory regarding race; it differentiates between normal children and learning disabled children; and its format is very appealing to young children (Paget, 1985).

However, the McCarthy is not without limitations. Many of the validity and reliability studies were based on small samples. The GCI is somewhat confusing and it doesn't compare well with intelligence quotients. Studies have found that the GCI can be very discrepant from IQ scores on the Stanford-Binet and WISC, by as much as one standard deviation (Woodrich, 1985). However, other reviewers indicate that the GCI is an index of intellectual functioning and can be used interchangeably with the IQ (Kaufman & Kaufman, 1977). It is also difficult to interpret scores derived from the memory and quantitative indexes; this should be done with caution (Paget, 1985). Another limitation of the McCarthy is that it lacks abstract reasoning and verbal expression components. Also, the McCarthy has limited floors and ceilings and consequently results in problems when testing younger children of low ability and highly gifted children (Kaufman & Kaufman, 1977; Woodrich, 1985). Also, language-delayed preschoolers are difficult to assess using this instrument. Woodrich states: "all in all, the MSCA represents one viable clinical tool for assessing preschoolers."

References:

Buros, O. K. (1978). *The Eight Mental Measurements Yearbook.* Highland Park, NJ: The Gryphon Press.

Kaufman, A. S. and Kaufman, N. L. (1977). *Clinical Evaluation of Young Children with the McCarthy Scales.* New York: Grune & Stratton.

Paget, K. D. and Bracken, B. A. (1983). *The Psycho-educational Assessment of Preschool Children.* New York: Grune & Stratton.

Paget, K. D. (1985). *Review of the McCarthy Scales of Children's Abilities.* In James V. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook.* Lincoln: The University of Nebraska Press.

Woodrich, D. L. (1985). *Review of the McCarthy Scales of Children's Abilities.* In James V. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook.* Lincoln: The University of Nebraska Press.

The Revised Milani-Comparetti Motor Development Screening Test

Author(s):	A. Milani-Comparetti & E.A. Gidoni Revised Edition-Wayne Stuberg
Publisher:	Meyer Rehabilitation Institute
Address:	University of Nebraska Medical Center
Copyright Date:	1992
Price:	$15

Purpose: The Milani is a series of simple procedures designed to determine whether a child's physical development corresponds to that of a normal child's.

Description: The Milani consists of procedures used to assess 27 motor behaviors (9 spontaneous behaviors and 18 evoked responses). The manual depicts a series of easy-to-learn procedures to screen motor development. The order of items from the original form has been changed to minimize handling and changing the position of the child. The authors report that although some behaviors may be observed, the test emphasizes the handling of the child to gain information on the select behaviors.

Range of Children: The procedures screen the motor development of children from birth to 2 years of age. The majority of test items are scored within the first 16 months of the child's life; however, the test provides the most detailed information regarding children between 3 to 12 months of age.

Testing Time: 10 to 15 minutes

Scoring: The recently revised chart for recording the child's responses is organized in grid fashion. Entries on the chart are made by writing the chronological age in months beneath the functional finding indicated at the head of the columns. No attempt is made to grade the child's responses; only their presence or absence is noted. Responses can also be coded as age appropriate, delayed, advanced, asymmetrical, not observed, or not tested. Items on the Revised Score Form are coded to represent the age when children in the normative sample demonstrated a particular reflex or motor behavior.

Examiner: Professional (usually an occupational therapist or physical therapist).

Standardization: The standardization sample is not representative of the U.S. population. Data on the normative sample were collected during 1985 and 1986. The sample consisted of 312 Omaha children (155 males and 157 females) from well-baby clinics, day care centers, and private homes between the ages of 1 and 16 months. Children were prescreened using the Denver Developmental Screening Test; only those who received a score of "normal" were included in the sample.

Reliability: Reliability of the Milani is adequate for a screening test. Two types of reliability were calculated: inter-observer and test-retest reliability. Interobserver reliability was calculated from the scored responses of three pediatric physical therapists who each watched the videotaped test administration for the same sample of 60 children. Agreement ranged from 89 percent to 95 percent. Test-retest agreement was calculated from the results of 43 children who were examined twice within a five- to seven-day interval by the same therapist. Test-retest agreement ranged from 82 percent to 100 percent.

Validity: Validity for the Milani is inadequate. Only content reliability is reported. The authors report, "content validity of the Milani test items relative to motor behavior has been recognized through its acceptance by physicians and therapists."

Discussion/Concerns: The psychometric information is incomplete. The standardization sample is limited to subjects taken from only one city. The sample is over-represented by infants from middle and upper class families, and minorities are underrepresented with regard to the U.S. population. In addition, validity is incomplete. Stuberg (1992) reports that although scoring systems have been developed, "the test is a 'screening' tool used to formulate an impression. The tool's validity in 'assessment' of motor handicaps has not been adequately demonstrated, and caution should be exercised in interpreting the results for diagnostic input or framing recommendations."

References:

Stuberg, W. (Ed.), *The Milani-Comparetti Motor Development Screening Test*. Omaha: Meyer Rehabilitation Institute, University of Nebraska Medical Center.

A Motor Development Checklist

Author(s): Anna M. Doudlah, Ph.D.
Publisher: Library Information Center
Address: Central Wisconsin Center for the Developmentally Disabled, 317 Knutson Drive, Madison, WI 53704
Copyright Date: 1976
Price: 1 Copy of Checklist and 5 scoresheets—$1, 50 scores sheets—$1, Videotape—$35, (specify reel-to-reel or 3/4" cassette). Short-term loan also available.

Purpose: The Motor Development Checklist was designed to pinpoint the level of a child's motor development from birth to walking in terms of spontaneous action patterns. Spontaneous action patterns are stated to be the most representative of a child's developmental status. The sequence of motor development described in the checklist can be used as a basis for planning and evaluating the effectiveness of motor development programs for children with developmental disabilities. The sequence of motor development is considered crucial, while time and rate of development are not as important.

Description: The checklist is an observational record of gross motor development from birth to walking. It consists of "Motor Development: Birth to Walking," an 18 minute videotape, five copies of "A Motor Development Checklist," and 25 scoresheets. A preview copy of the videotape is available on short-term loan.

Range of Children: Normal and developmentally disabled children. Approximate ages: 1-15 months (normal development) or during time period of birth to walking.

Testing Time: Observation and recording to be done monthly. Length of observation is dependent on spontaneous motor movement of the child.

Scoring: The checklist can be used in the following ways: (1) simply as a checklist by placing a mark in the appropriate square, noting which motor behaviors were seen during the observation period or (2) utilizing the following scale during each observation period.

0 - Does not perform task
1 - Beginning to attempt task
2 - Performs task occasionally
3 - Performs task skillfully

The latter method provides more time-related information about the child's developmental progress. Observation and recording should be done monthly and the date of the observation is noted in the square provided for that purpose. Spaces are provided for 24 observation periods on the score sheet.

Examiner: Professional, usually Physical or Occupational Therapist.

Standardization: The checklist derived from the movie records of the spontaneous motor behavior of 20 normal infants filmed monthly in their own homes over a 15-month period. The checklist is the result of a longitudinal study and is not norm-referenced.

Reliability: No data reviewed.

Validity: No data reviewed.

Discussion/Concerns: The concept of the acquisition of gross motor skills that are self-induced or spontaneous is valid in terms of obtaining an accurate assessment of a child's gross motor status. This is the basic concept underlying Doudlah's Motor Development Checklist that makes it different from other measures of motor development. By basing a gross motor program on the observed spontaneous action patterns of a child, there is less risk of beginning a program at the wrong point in a child's development, and less chance that specific motor movements will be omitted.

An interesting aspect of Doudlah's checklist is the inclusion of an item called "oscillation." This has not been included in other schedules or checklists of motor development. Doudlah defines oscillation as a motor movement in place that prepares or stabilizes the muscles for further development of movement through space. Oscillation is a temporary phenomenon that disappears from the child's movement patterns after muscles are stable enough for moving in various ways.

Doudlah's Motor Development Checklist would be a useful tool for teachers in communicating with occupational and physical therapists concerning gross motor programs. It could also be used as a guide for parents for understanding the sequence of their child's motor development.

The Motor Development Checklist is easy to use in terms of administration, and the cost is reasonable.

References:

Doudlah, A. (1978). Early motor development: intrauterine to walking. In *Encyclopedia of Physical Education and Sports*. Redding, MA: Addison-Wesley.

Piker, E. (1971). Learning of motor skills on the basis of self-induced movements. In J. Hellmuth (Ed.), *Exceptional Infant, Studies in Abnormalities*. New York: Brunner/Mazel.

Motor-Free Visual Perception Test (MVPT)

Author(s):	Ronald Colarusso & Donald Hammill
Publisher:	Academic Therapy Publication
Address:	20 Commercial Blvd., Novato, CA 94949
Copyright Date:	1972
Price:	$55 per kit

Purpose: The MVPT is a test of visual perception which avoids motor involvement and which is practical for screening, diagnostic, and research purposes (Calurros & Hammill, 1972).

Description: The MVPT is a 36-item test which is individually administered. It measures five categories of visual perception: spatial relationships, visual discrimination, figure ground, visual closure, and visual memory. The child is shown an abstract figure drawing and is asked to point to one of the four alternatives which he/she thinks is the correct response. Children are given the entire test.

Range of Children: The MVPT can be administered to children from ages 4-0 to 8-11.

Testing Time: Test administration takes approximately 10-15 minutes.

Scoring: The child's answers are recorded on an accompanying scoring sheet. Raw scores can be converted to perceptual ages and perceptual quotients ($\underline{M}$ = 100, $\underline{SD}$ = 15).

Examiner: Professional, usually an education specialist or psychologist.

Standardization: The MVPT was standardized on a random sample of 881 normal children ages 4 through 8. Sample subjects were from 22 states and from all races, economic levels, and residential areas. Children who were identified as mentally retarded, sensorially handicapped, etc., were excluded.

Reliability: Test-retest reliability ranged from .77 to .83 at different age levels, with an average of .81 for the entire sample. Few 4-year-olds were included in the sample; therefore, test results of this age group should be interpreted with caution (Calarusso & Hammill).

Validity: Authors of this instrument indicate that content validity is adequate. They also contend that construct validity is adequate as scores improved with age in a study of 40 children (Calarusso & Hammill). Concurrent validity was assessed by comparing the MVPT to several other visual perception tests. Correlation coefficients ranged from .31 to .73 with a mean of .49. The MVPT correlated with the Frostig Test at .73. The MVPT was also compared to several achievement, readiness, and intelligence tests.

Correlation coefficients varied, with the highest correlation being .50 between the MVPT and Metropolitan Readiness Tests.

Discussions/Concerns: The MVPT is a fairly weak instrument psychometrically. Reliability is low and is generally considered inadequate, especially for diagnostic purposes. Concurrent validity is also inadequate. Also, the fact that no special needs children were included in the norm sample is considered a weakness. Strengths of this test are that it is visually appealing to children and may effectively tap into visual-perceptual problems. This test could be effective as a screening instrument but should be used with caution as a diagnostic tool.

References:

Calarusso, R. & Hammill, D. (1972). *MVPT Manual: Motor-Free Visual Perception Test*. Novato, CA: Academic Therapy Publications.

Ordinal Scales of Psychological Development

thor(s):	Ina C. Uzgiris & J. McV. Hunt
Publisher:	University of Illinois Press
Address:	54 E. Gregory Dr., Champaign, IL 61820
Copyright Date:	1975, 1989
Price:	$12.95 - manual (1989 paperback edition) $15 - record forms (five)

Purpose:

The six ordinal scales investigate the effects of infants' encounters with various kinds of circumstances in relation to cognitive development.

Description:

The Ordinal Scales were originally developed as a research instrument and were based on Piagetian principles suggesting that intellectual development in infancy is sequential, hierarchical, and invariant. The authors identified a number of actions that reflect varying levels of cognitive structure and organization. These "critical actions" are observed when the infant is confronted with a series of "eliciting situations." According to Uzgiris and Hunt, the specific assessment procedures and materials are less important than is a theoretical understanding of what the assessment instrument is; thus, it is highly theoretical and dependent upon both the observational and interpretive skills of the evaluator.

There are six individually administered scales: (1) Development of Visual Pursuit and the Performance of Objects, (2) Development of Means for Obtaining Desired Environmental Events, (3) Development of Vocal Imitation and Gestural Imitation, (4) Development of Operational Causality, (5) Construction of Object Relations in Space, and (6) Development of Schemes for Relating to Objects. It is not necessary to present the scales in a single session or in sequence, nor is it necessary to administer all of the scales. It is necessary, however, to present situations appropriate for eliciting the critical actions for several consecutive steps on the scale to ascertain the infant's level of development.

Development and administration of the Ordinal Scales is outlined in the book *Assessment in Infancy: Ordinal Scales of Psychological Development*. A supplemental manual ($24) written by Carl Dunst and published by Pro-Ed is titled *A Clinical and Educational Manual for use with the Uzgiris and Hunt Scales of Infant Psychological Development*. Both are recommended. The Dunst manual contains helpful, elaborative information for administration as well as record forms which are easy to use; however, Dunst does not repeat the basic administrative procedures which are outlined in the Uzgiris and Hunt book. Dunst provides case material demonstrating the use of the scales and guidelines for developing interventions based on these scales.

In addition to the two manuals, the administrator will need to compile a set of materials to use in the eliciting situations. Many of these can be pulled from other tests or from the preschool toy box. All are easily obtained.

Range of Children: The populations examined in developing the scales ranged in age from 1 to 24 months. The scales have also proved useful in assessing older severely handicapped individuals (see Uzgiris & Hunt, 1986, and Dunst, 1980, for details).

Testing Time: The amount of time necessary will depend upon the age and motivation of the child and the number of scales utilized by the examiner.

Scoring: The scales are intended to provide the examiner with qualitative information rather than numerical scores. This instrument does not rely upon comparison scores with other children or a standardization sample for interpretation. The authors suggested that the raw scores be used to determine the kind of circumstances required to promote succeeding steps in development.

Examiner: Persons well trained in observation of young children and in the theory underlying the scales.

Standardization: No attempts to obtain normative data have been carried out by the authors. In development of the scales, however, 149 infants ranging in age from 1 to 24 months were observed.

Reliability: Three pairs of examiners observed 84 children to establish inter-examiner reliability and test-retest reliability. Interobserver reliabilities varied across age groups and eliciting situations from a low of .42 to high of 1.00. Most reliabilities were between .85 and 1.00, however. Across all actions, the mean test-retest coefficient was .799.

Validity: Uzgiris and Hunt utilized a five-step process in developing the instrument. First, various actions of infants described by Piaget as indicative of new levels of cognitive organization were designated. Second, the infant actions and the situations Piaget had used to elicit them were arranged into a schedule and instructions were prepared for eliciting as well as observing and recording the behaviors. Third, a sample of infants representing every month of age from birth to two years was observed in homes. Fourth, a second examiner was trained to determine test-retest reliability. Fifth, the instrument was used in a longitudinal study to establish that the hypothesized sequential order was indeed invariant and persisted through different systems of child rearing. Additional studies supporting the invariant, sequential nature of development are reported in Uzgiris and Hunt (1986).

Discussion/Concerns:

The Ordinal Scales are based on an approach to assessing very young children that differs substantially from other traditionally used instruments. The items largely address the degree to which the child responds to and interacts with stimuli in his or her environment. The examiner is able to assess such things as the desire for continued contact with an object no longer visible, the means employed to cause desired events or obtain desired objects, and the ability to elicit imitation behaviors. Also assessed are understanding of causality, the ability to determine the positions objects occupy in space, and the ways in which the child interacts with common toys. Uzgiris and Hunt make few claims regarding far-reaching generalizability or predictive implications. As a qualitative instrument, the Ordinal Scales have no real competition. They are useful as a means of understanding development more than quantifying it.

References:

Dunst, C. J. (1980). *A Clinical and Educational Manual for use the Uzgiris and Hunt Scales of Infant Psychological Development*. Austin, TX: Pro-Ed.

Uzgiris, I. C. & Hunt, J. Mc.V. (1986). *Infant Performance and Experience: New Findings with the Ordinal Scales*. Urbana: University of Illinois Press.

Uzgiris, I. C. & Hunt, J. Mc.V. (1989). *Assessment in Infancy: Ordinal Scales of Infant Psychological Development*. Urbana: University of Illinois Press.

Peabody Developmental Motor Scales & Activity Cards (PDMS)

Author(s):	M. Rhonda Folio & Rebecca R. Fewell
Publisher:	DLM Teaching Resources
Address:	One DLM Park, Allen, TX 75002
Copyright Date:	1983
Price:	$195 per complete program

Purpose: The PDMS is an early childhood motor development program that provides both in-depth assessment and instructional programming for gross and fine motor skills.

Description: The PDMS tests several domains including reflexes, balance, nonlocomotor, locomotor, receipt and propulsion of objects, grasping, hand use, eye-hand coordination, and finger dexterity. The test is administered individually and involves having the child perform various tasks. Some materials are included in the test; however, some local assembly is required (balance beam, pull toy, etc). Activity cards which outline program and training procedures for remediation purposes are also provided (Reed, 1985).

Range of Children: The PDMS can be used to assess children from birth to 83 months.

Testing Time: Administration time is approximately 20 to 30 minutes for each scale.

Scoring: The child's performance on test items is recorded on a scoring sheet. Basal and ceiling levels are provided. Raw scores can be converted to scaled scores (Z-scores, T-scores, developmental motor quotients) and age equivalents.

Examiner: Professional, usually an occupational therapist.

Standardization: The PDMS was standardized on 617 children. Distribution of Caucasian, African-American, and Hispanic categories parallel U.S. census data.

Reliability: According to a review by Reed (1985), reliability of each scale is "well established."

Validity: No information available.

BEST COPY AVAILABLE

Discussion/Concerns: Reed (1985) concludes that the psychometric properties of this test are somewhat weak. The norm group is limited and does not include adequate numbers at each age level. Reed points out that the activity cards included in the PDMS received no critical analysis in the test manual. A strength of the PDMS is that it attempts to quantify children's motor development where "eyeball" judgements have predominantly been used. Scoring of the gross motor and fine motor scales is clearly objective. Also, the PDMS is directly responsive to PL 94-142. In summary, Reed says, "PDMS is a positive contribution that will surely find wide applicability."

References:

Reed, H. B. C. (1985). Review of Peabody Development Motor Scales and Activity Cards. J. T. Mitchell (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Peabody Picture Vocabulary Test (PPVT-R)

Author(s):	Dunn, Loyd M. and Dunn, Leota M.
Publisher:	American Guidance Service
Address:	Box 99, Circle Pines, MN 55014
Copyright Date:	1981
Price:	$113.50 for Kit (Forms L and M)

Purpose: The PPVT-R is designed to measure a subject's receptive vocabulary for Standard American English.

Description: The PPVT-R is an individually administered instrument that contains two forms: Form L and Form M. The instrument contains 175 test items for each form and there are five training plates to be administered to subjects. Each test plate contains four pictures from which the subject is to choose the picture that best describes/defines the word said by the examiner. The test record form lists the stimulus words to be used with the training and test plates.

Range of Children: The age range for the PPVT-R is from 2½ to adult.

Testing Time: The authors recommended administration time is 10 to 20 minutes.

Scoring: The subject's total raw score on the PPVT-R is the number of correct responses over the critical range (the range between the test basal and test ceiling). Starting points for determining the test basal are provided (all items above the ceiling are counted incorrect and all items below the basal are counted as correct). The raw score can be converted to a percentile rank, age equivalent score, or to a standard score with a mean of 100 and a standard deviation of 15.

Examiner: The examiner should be thoroughly familiar with the test materials. Interpretation of test results is to be done by a trained professional.

Standardization: The PPVT-R was standardized on a representative national sample of children and youth, and a selected sample of adults. The norm sample for ages 2½ to 18 years included 4,200 children. The sample was stratified by age, sex, geographic region, occupation, ethnic background, community size, and by the two test forms (Form L and Form M given about equally).

Reliability: Split-half correlations based on all subjects in the standardization sample (ages 2½ through 40) were obtained. For children and youth (2½ through 18) the coefficients ranged from .67 to .88 on Form L (median .80) and from .61 to . 86 on Form M (median .81). For PPVT-R standard scores, the reliabilities range from .71 to .89 with a median of .79 on immediate retest/alternate forms reliability.

Delayed retest/alternate forms reliability coefficients for standard scores ranged from .54 to .90 with a median of .77. The authors indicate that the reliability of the PPVT-R appears to be satisfactory. Also, they state that, "Generally, the revised PPVT is a slightly more reliable measure than the original PPVT."

Validity:

No PPVT-R validity information of a statistical nature is provided in the manual. Instead, the authors report a study equating the PPVT and PPVT-R as evidence of validity. The authors indicate the PPVT-R has content validity and construct validity (as long as it is seen as a measure of hearing vocabulary). Independent receivers collaborate this conclusion.

Discussion/Concerns:

The PPVT-R is designed to measure a subject's receptive vocabulary for Standard American English. The authors indicate that it can provide a quick estimate of one major aspect of verbal ability. It is *not* a comprehensive test of general intelligence. The PPVT-R is easy and quick to administer. Scores are easily converted to derived scores such as age, standard, or percentile measures. The materials are durable and interesting to subjects.

Important new features of the PPVT-R are (1) the terms *mental age* and *intelligence quotient* were changed to *age equivalent* and *standard score equivalent*, (2) adult norms were added, and (3) standardization was conducted on a national sample. The PPVT-R is such a significant improvement over the PPVT that the earlier version should not be used for educational decision making. Eliminating the term *intelligence quotient* was a commendable step since many users of the PPVT inappropriately used this score as a measure of global intelligence. The PPVT-R is not an intelligence test; it is an estimate of receptive vocabulary.

References:

Dunn, L. M. and Dunn, L. M. (1981). *Peabody Picture Vocabulary Test: Revised—Test Manual*. Circle Pines, Minnesota: American Guidance Service.

Mitchell, J. V. Jr. (1985). *Ninth Mental Measurements Yearbook*. Lincoln: University of Nebraska Press.

Personality Inventory for Children (PIC)

Author(s): Robert D. Wirt, Daniel Lachar, James K. Klinedinst, & Philip D. Seat
Publisher: Western Psychological Services
Address: 12031 Wilshire Boulevard, Los Angeles, CA 90025
Copyright Date: 1982
Price: $210 per kit

Purpose:

The Personality Inventory for Children (PIC) is an instrument which seeks to provide a comprehensive and clinically relevant personality description of children.

Description:

The PIC booklet contains 600 questionnaire items which can be answered true or false dependent on the respondent's opinion of a child's behavior, attitudes, and family relationships. Not all 600 items need to be administered. The child's mother is recommended as the primary respondent as this instrument was normed on this group. The PIC contains 3 validity, 1 general screening, 12 clinical, and 17 supplemental experimental scales. The validity scales are used to determine if the respondent is providing valid responses, and the clinical scales seek to identify children who may need further psychological evaluation as well as to serve as a general measure of psychological maladjustment.

Materials for the PIC include a booklet and answer sheet used by the respondent and a manual for interpretation and scoring purposes. There are templates for hand scoring, or computer programs for scoring are available.

Range of Children:

The PIC can be used to evaluate both male and female children ranging in age from 3 to 16 years.

Testing Time:

Dependent on which scales are given: 45 minutes to two hours.

Scoring:

True/false answers are recorded onto computerized answer sheets. There are two separate PIC profile sheets: one for ages 3-5 and the other for ages 6-16. Raw scores are converted to standard scores (T-scores) for interpretative purposes.

Examiner:

Professional psychologist or psychiatrist must interpret the PIC.

Standardization:

The 1982 PIC was not restandardized in any way and the original norms continue to be used. The PIC was standardized in 1960 on 2,390 children in the Minneapolis Public School System. Several hundred additional cases were collected in 1970. In 1977, the authors collected records for subjects from age 2-6 to 5-6 (102 boys and 90 girls). The authors indicate that the families used for standardization had a good distribution of economic, social, and educational backgrounds.

Reliability: Test-retest reliability ranged from .46 to .94 with a mean reliability of .86 for psychiatric outpatient subjects. In two different studies of normal children, test-retest reliability ranged from .50 to .87 (M .71) and from .68 to .97 (M .89). Internal consistency ranged from .57 to .86 with a mean of .74.

Validity: Studies which have been conducted addressing this instrument's concurrent, convergent, and discriminate validity have created an excellent foundation for the PIC. However, it does require further work to be considered adequate (Knoff, 1989).

Discussion/Concerns: The PIC is comprehensive and is considered to have a theoretical background similar to that of the Minnesota Multiphasic Personality Inventory. This instrument is psychometrically weak and needs to be restandardized and appropriately stratified (Knoff, 1989). According to Knoff (1989) norms are dated and geographically localized, and pre-school norms are extremely limited. However, it has a good research foundation supporting its psychometric potential (Knoff, 1989). The reorganization of test items is also a strength of this instrument. The PIC appears to be a good screening diagnostic instrument as compared to other similar assessment tools (Knoff). It can provide important and valuable information concerning a child's personality development. However, its use with preschoolers is questionable due to scores being based on a small norm group.

References:

Knoff, H. M. (1989). Review of the Personality Inventory for Children. In J. Close Conoley & J. K. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Lachar, D. & Golowski, C. L. (1979). *Actuarial Assessment of Child and Adolescent Personality: An Interpretive Guide for the Personality Inventory for Children Profile*. Los Angeles: Western Psychological Services.

Wirt, R. D. (1981). *Multidimensional Description of Child Personality—A Manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.

Photo Articulation Test (PAT)

Author(s):	Kathleen Pendergast, Stanley Dickey, John Selmar, and Anton Soder
Publisher:	Pro-Ed
Address:	8700 Shoal Creek Blvd, Austin, TX 78758
Copyright Date:	1984
Price:	$69 per kit

Purpose: The PAT enables the speech and language clinician to assess a child's articulation using a series of photographs.

Description: The test consists of 72 colored photographs (nine photos to a page). The first 69 photos test consonants and all but one vowel and one diphthong. The last three pictures assess connected speech and the remaining vowel and diphthong. The child is shown photographs that are designed to elicit the target response. A deck of the same photos is also available for children who have trouble with more than one picture to a page. A supplemental test words list is also included to further assess articulation errors (Pendergast et al, 1984).

Range of Children: Children ranging in age from 3-0 to 11-11 can be assessed using the PAT.

Testing Time: Test administration ranges from 5 to 10 minutes.

Scoring: After the child responds to the picture, the examiner records the response on a sheet which is divided to provide separate scores for three categories: tongue sounds, lip sounds, and vowel sounds. Recording responses takes into account if the child pronounces the word accurately, omits sound, substitutes sound, and distorts sound. Articulation Age Overlays (AAO) which show age-appropriate sounds are provided. The AAOs are transparencies which can be placed over the recording sheets to help the examiner visually compare the subject's articulation errors with norms.

Examiner: Professional, usually a speech and language clinician.

Standardization: Norms were developed using a sample of 684 Caucasian children ranging in age from 3-12 years. The children were randomly selected and were from Seattle, WA. They were generally from middle socio-economic homes.

Reliability: Test-retest reliability was assessed using a sample of 100 children. A correlation coefficient of .99 was found. No internal consistency or interrater reliabilities were reported in the manual.

BEST COPY AVAILABLE

Validity: Concurrent validity between the PAT and two other tests was assessed using a sample of 100 children. A correlation coefficient of .97 was found between the PAT and Bryngelson-Glaspey. The Templin-Darley correlated with the PAT at .82. Content and construct validity were not discussed in the manual.

Discussion/Concerns: Validity and reliability of this test have not been effectively assessed and the studies which have been done have used limited sample data. Consequently, the psychometric properties of this test appear inadequate. Also, children with articulation problems were not used in norming this instrument; consequently, caution should be used when assessing this population using the PAT. In general, the PAT appears to yield valuable information concerning patterns of specific articulation errors. The photos are colorful and interesting and testing time is short. However, further development of the PAT's psychometric properties is needed.

References:

Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1984). *Photo Articulation Test*. Austin, TX: Pro-Ed.

97

Portage Guide to Early Education (Revised)

Author(s): S. Bluma, M. Shearer, A. Frohman, J. Hilliard
Publisher: The Portage Project
Address: Cooperative Educational Service Agency 12, 412 East Slifer Street, Portage, WI 53901
Copyright Date: 1976
Price: Contact publisher

Purpose: The *Portage Guide to Early Education* (revised edition) was developed to serve as a guide to those who need to assess a child's behavior and plan realistic curriculum goals that lead to additional skills. The checklist and card file can aid in assessing present behavior, targeting emerging behavior, and providing suggested techniques to teach each behavior.

Description: The Guide contains three parts: (1) a checklist of behaviors on which to record an individual child's developmental progress, (2) a card file listing possible methods of teaching these behaviors, and (3) a manual of directions for use of the checklist and card file as well as methods for implementing activities.

The checklist serves as a method of informal assessment. The checklist is color coded and divided into six developmental areas: Infant Stimulation, Socialization, Language, Self-Help, Cognitive, and Motor. A checklist can be completed on each child upon entry into a program. The checklist can serve as an ongoing curriculum record for all of the preschool years; essentially, the same checklist can be used each year.

The behaviors are listed sequentially, at one-year intervals, in each category from birth to 6 years. The Guide is designed to be a curriculum planning tool. The information derived from its use is utilized to delineate those skills acquired and those yet to be taught.

The skills listed on the checklist are behaviorally stated. No specific criteria are provided, although some items do include examples. The examiner might refer to the card file to determine specific activities that could be used to assess the skill. There is a total of 580 items, 535 if the Infant Stimulation items are not utilized.

Range of Children: The checklist as well as the entire *Portage Guide to Early Education* can be used with children between the mental ages of birth and six years of age; the materials can be used with normal preschool children or preschool children with handicaps.

Testing Time: No information reported.

Scoring: To the right of the "Behavior" column (which designates the individual skills) are three columns that are utilized for recording purposes. If the child demonstrates the skill at the time he/she enters the program a check (✓) is placed in the "Entry Behavior" column. If the skill is not demonstrated at that time, the date of achievement is recorded in the "Date Achieved" column at such time that the child acquires the skill. The third column is for comments regarding characteristics observed while assessing the individual skills or gathering information via parent report. No procedures for obtaining standard scores are provided.

Examiner: Professional and paraprofessional (teachers, aides, nurses, parents, others).

Standardization: No information reported.

Reliability: No information reported.

Validity: No information reported. In discussing modifications made in the revised form, the authors state, "All the behaviors which formerly appeared in the checklist have been reevaluated as to their importance in child development and consequently some have been deleted. Other behaviors have been added." The bibliography lists such authorities as Alpern and Boll, N. Bayley, Roger Brown, P. Cattell, E. Doll, L. Dunn, W. K. Frankenburg et al., A. Gessell, M. Sheridan, R. Slossen, L. Terman and M. Merrill, D. Wechsler, and Burton White. It is expected that many of the items involve skills similar to those delineated by such authorities.

Discussion/Concerns: The *Portage Guide to Early Education* serves as a resource to professionals and nonprofessionals who are involved in the delivery of services to young children. The checklist may serve as an adequate assessment guide when serving nonhandicapped populations. However, additional formal procedures are warranted when serving handicapped children. In fact, the checklist might be utilized following formal assessment procedures to determine some specific instructional goals and procedures.

References: Bluma, S., Shearer, M., Frohman, A. and Hilliard, J. (1976). *The Portage Guide to Early Education: Revised*. Portage, WI: Cooperative Educational Service Agency #12.

Preschool Language Scale—3 (PLS-3)

Author(s):	Irla Lee Zimmerman, Violet G. Steiner, Roberta Evatt Pond
Publisher:	The Psychological Corporation
Address:	PO Box 839954, San Antonio, TX 87283-3954
Copyright Date:	1991
Price:	$89 (Complete Kit)

Purpose: The PLS—3 is a norm-referenced measure designed to assess young children in the areas of receptive and expressive language for eligibility for special education and for program planning. A Spanish version is also available.

Description: The PSL—3 uses toys and pictures to engage children in pre-language and language activities. Activities are developmentally appropriate and naturalistic. The test has 48 receptive and 48 expressive items. The test measures development on four aspects of language including language precursors, semantics, structure, and integrative thinking skills. The test includes a parent input component, language sample form and checklist, and an articulation screener. Modifications for children with hearing impairments or physical disabilities are suggested.

Range of Children: The PLS—3 can be used with children birth through 6 years of age.

Testing Time: No testing time indicated

Scoring: A child's score is recorded as a standard score, in percentile rank and as an age-equivalent score. Separate scores are obtained for auditory comprehension and expressive communication as well as a total language score.

Examiner: Professional

Standardization: This instrument was normed on more than 1,800 children across the U.S. Additional information is available in the test manual.

Reliability: Data are reported in the test manual.

Validity: Data are reported in the test manual.

Discussion/Concerns: The PLS—3 has overcome some of the problems with the earlier PLS—R. The inclusion of more objects and naturalistic activities (including observations of spontaneous behavior) in the test items makes the test more functionally oriented in its administration and interpretation. The test has a brief administration time and is easily administered and recorded. It addresses varied expressive and receptive tasks using manipulative and colored pictures. Tasks target social/interactive communication skills and integrative thinking skills.

Articulation is considered separately from expressive language development. Family information is an integral part of the evaluation.

A problem throughout the test is its lack of test items for certain areas at different levels. For example, if a child reached a basal at item #31 and a ceiling at item #38 on the Auditory Comprehension section, the areas of quantity and morphology would not have been tested. On the expressive communication section there is only one item for each of the areas of quantity, quality, and spatial relations. Performance on one item is not sufficient to make a judgement. The PLS—3 serves well as a test to be used first in a battery of tests as determiner of which areas need further assessment.

The supplemental articulation screener appears to be developmentally appropriate. The imitative method of elicitation may result in underidentification. The speech intelligibility rating scale (good, fair, poor) needs to be supplemented with descriptive data.

References:

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool Language Scale—3*. San Antonio, TX: Psychological Corporation.

Strand, L. & Jacobs, E. (1992). Evaluation of the PLS—3. Personal communication with the Arizona Department of Education/Special Education Section/Preschool Unit, Northern Arizona University, Institute for Human Development, Flagstaff.

Scales of Independent Behavior (SIB)
(Woodcock-Johnson Psycho-Educational Battery—Part 4)

Author(s): Robert H. Bruininks, Richard W. Woodcock, Richard F. Weatherman, and Bradley K. Hill

Publisher: DLM Teaching Resources

Address: One DLM Park, P.O. Box 4000, Allen, TX 75002

Copyright Date: 1984

Price: $135 for Complete SIB Program

Purpose: The SIB is designed to assess behaviors needed to function independently in home, social, and community settings.

Description: The test is administered to an informant who rates the behavior of the child on a number of items. The test consists of four adaptive behavior clusters which include two to five subscales. They are as follows:

1. Motor Skills Cluster: Gross Motor, Fine Motor.

2. Social and Communication Skills Cluster: Social Interaction, Language Comprehension, and Language Expression.

3. Personal Living Skills Cluster: Dating and Meal Preparation, Toileting, Dressing, Personal Self-Care, and Domestic Skills.

4. Community Living Skills Cluster: Time and Punctuality, Money and Value, Work Skills, and Home-Community Orientation.

The Broad Independence Cluster (Full Scale) is a measure of independence based on the results of all 14 subscales. A Short Form Scale provides a brief measure of broad independence and the Early Development Scale provides a developmental measure of adaptive behavior from infancy to three years; it is also useful for assessing the functioning of severely and profoundly handicapped individuals.

Also included in the SIB is a scale for assessing eight areas of problem behavior including: Hurtful to Self, Hurtful to Others, Destructive to Property, Disruptive Behavior, Unusual or Repetitive Habits, Socially Offensive Behavior, Withdrawal or Inattentive Behavior, and Uncooperative Behavior. Special scores include four indexes of maladaptive behavior: the Internalized Maladaptive Index (IMI); the Asocial Maladaptive Index (AMI); the Externalized Maladaptive Index (EMI); and the General Maladaptive Index (GMI).

Range of Children: Norms are provided for infants to adults (over 40 years). The Early Development Scale allows the test to be used with infants and severely retarded individuals developmentally 2½ years or younger.

Testing Time: The manual recommends allowing one hour for administering the full scale and 15 minutes for the other options.

Scoring: The adaptive behavior section of the SIB utilizes a four point rating scale (0-3). The scale ranges from Never or Rarely Does a Task to Does Task Very Well without Help or Supervision. A raw score is used to determine cluster scores, age scores, instructional range, percentile rank, Relative Performance Index (RPI), Broad Independence score, Standard Scores and Normal Curve Equivalents. The Problem Behavior Scale utilizes two rating scales. The frequency of the problem behavior is rated on a six point scale (0-5) ranging from Never to One or More Times per Hour. Severity of the behavior is rated on a five point scale (0-4) ranging from Not a Problem to Critical Problem. The raw scores can be converted to stanines.

Examiner: Psychologist or Special education teacher.

Standardization: All norming data were obtained from individual interviewer-respondent protocols based on the complete SIB battery. The standardization sample for the SIB included more than 1700 subjects drawn from 39 communities varying in size, geographic location, and socioeconomic characteristics. Normative data were gathered from infancy through mature adult levels. Additional technical data were obtained on approximately 1000 handicapped and nonhandicapped subjects.

Reliability: Internal consistency is excellent for the Broad Independence score (.95 to .97), good for cluster scores (.83 to .93) and highly variable for the subscale scores (Heifetz, 1989). Internal consistency for the short forms is inadequate.

Test-retest reliability for the Broad Independence score was .87 to .96 and .71 to .96 for the cluster scores. However, only 6 to 10 and 10 to 11-year-olds were used in the study (Heifetz, 1989). Interrater reliability based on a sample of retarded adolescents ranged from .74 to .86 for the adaptive behavior scale and .69 to .81 for the problem behavior scale.

Validity: According to the authors, the SIB possesses good content validity in that it assesses the broad range of skills and traits included in models of adaptive and maladaptive behavior. Also, the authors indicate that the SIB does differentiate between handicapped and nonhandicapped individuals. Reviewers disagree concerning the adequacy of content validity as discussed in the manual (Heifetz, 1989; Camp, 1989). Concurrent validity is considered adequate since the SIB was correlated with several tests which yielded high correlation coefficients (Heifetz, 1989). No evidence of predictive validity was discussed in the manual.

Discussion/Concerns:

The SIB is easily learned, employs precise behavioral statements, and provides useful information about the adaptive functioning of normal and handicapped children (Camp, 1989). However, there is some concern about how items are rated since they fail to take into account the capability of the person versus his/her motivation level. That is, a person may be able do the task, but may choose not to (Heifetz, 1989). A strength of the SIB is that it includes a scale for behavior problems.

The SIB is fairly easy to administer; however, the easel is somewhat cumbersome and scoring can be time consuming. The authors recommend that it be used only in conjunction with the rest of the Woodcock-Johnson (WJ-R) battery and caution that comparisons made between the SIB and other intelligence measures may not be valid. Since the norm groups were not the same for the SIB and the WJ-R, however, this caution is spurious. Some of the validity and reliability data appear adequately assessed; however, other aspects of reliability and validity are ignored altogether (Heifetz, 1989). Heifetz concludes that, overall, "in comparative terms, the SIB is psychometrically one of the better indirect-report measures of adaptive behavior."

References:

Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1984). *Manual-Scales of Independent Behavior*. Allen, TX: DLM Teaching Resources.

Camp, B. W. (1989). Review of the Scales of Independent Behavior. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Heifetz, L. J. (1989). Review of the Scales of Independent Behavior. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Sequenced Inventory of Communication Development—Revised Edition (SICD-R)

Author(s): Dona Lea Hedrick, Elizabeth M. Prather, and Annette R. Tobin
Publisher: University of Washington Press
Address: Seattle, WA 98195
Copyright Date: 1984
Price: $250 per complete kit. Some local assembly required.

Purpose: This test functions as a diagnostic assessment to evaluate the communication abilities of normal and handicapped children. It screens a breadth of early communication skills including semantic, syntactic, and pragmatic aspects of expressive and receptive language.

Description: This test is divided into two different scales: receptive and expressive. The receptive scale yields three scores which include awareness, discrimination, and understanding. The expressive scale yields four scores which are imitating, initiating, responding, and verbal output. The child's answers on these scales can be transferred to a Behavioral Profile or to a Processing Profile or both. Transfer of scores to these scales can be cumbersome, but provides valuable information.

The test consists of 100 items which are generally appealing to children. Six of the items must be provided by the examiner using paper, coins, and picture books. The test format involves varied response requirements such as object manipulations, following commands involving objects, picture identification, etc. The test includes a 50-item language sample for children over 2 years of age and an additional articulation test is recommended, but not provided. Also, parent report forms and observation of communication behavior forms are included. Furthermore, the authors include some valuable information for testing children with autism, hearing impairments, children who are difficult to test, and children of Eskimo heritage. A Spanish version is also included, but without norms.

Range of Children: The test includes a developmental range of 4 to 48 months.

Testing Time: Testing time varies, but usually lasts 30 to 75 minutes.

Scoring: The manual provides clear directions for administration, scoring, and computations of basals, ceilings, developmental ages, and Mean Length Responses. During administration, binary scoring is used.

Examiner: The instrument was designed to be administered by a speech and language clinician.

Standardization: The 1975 norms were based on 252 children from Seattle and were supplemented with data from 609 children from Detroit. The sample size is considered small and not representative of the U.S. population (Mardell-Czudnowski, 1989). Also, percentiles are not provided, and there are some developmental and chronological ages which do not match. An attempt was made to include African-American children in the sample; however, their ages ranged from 31-48 months which does not adequately reflect age levels of the entire population (Mardell-Czudnowski, 1989).

Reliability: Test-retest reliability and interrater reliability are acceptable; however, a small sample was used in the reliability research. Consequently, reliability of this test has not been sufficiently established (Mardell-Czudnowski, 1989). No standard error of measurements is reported.

Validity: Content validity is "supported by the test's reliance on both behavioral and language processing models" according to Mardell-Czudnowski. However, at some age levels few items were included. Also, inadequate information is given concerning construct and predictive validity. Concurrent validity of the SICD-R has not been established.

Discussion/Concerns: The SICD-R includes interesting, colorful materials which are appealing to young children. The two different profiles which can be obtained by using this instrument are helpful, and it does provide a good assessment of a wide range of communication behaviors. It also makes an attempt to include information on special populations and provides a Spanish translation. Some racial norms are provided, but do not span the entire age range. Some content validity is evident; however, overall reliability, validity, and standardization procedures are not considered adequate (Mardell-Czudnowski, 1989). Clear cutoff scores and evidence of the validity of the SICD-R as a screening instrument are not provided (Pearson, 1989). Although the SICD-R is an improvement over the original test, it appears to be an inadequate assessment tool; however, after a child has been assessed as delayed, the SICD-R could be useful in program planning (Mardell-Czudnowski, 1989).

References:

Mardell-Czudnowski, C. (1989). Review of the Sequenced Inventory of Communication Development—Revised Edition. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Pearson, M. E. (1989). Review of the Sequenced Inventory of Communication Development—Revised Edition. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Skills Inventory (The Oregon Project for Visually Impaired and Blind Preschool Children)

Author(s): Donnise Brown, Vickie Simmons, Judy Methvin
Publisher: Jackson County Education Service District
Address: 101 North Grape Street, Medford, OR 97501
Copyright Date: 1978
Price: $50

Purpose: The purpose of the Skills Inventory is to provide assessment and curriculum guidance to educators of young children with visual deficits.

Description: The Inventory has three components: The Manual, the Skills Inventory, and Teaching Activities. The Skills Inventory serves three purposes: (1) to assess the child's developmental level in six categories, (2) to select appropriate teaching goals, and (3) to record the child's acquisition of new skills. The Skills Inventory is not a tool for determining a precise score and should not be considered a normative assessment instrument.

The Skills Inventory assesses the child's development in the areas of cognition, language, self-help, socialization, fine and gross motor skills. The skills are organized by one-year intervals. A total of 693 skills are assessed. Appropriations of items for several levels of visual deficits are also specified. The items are presented in behavioral terms and are generally clearly stated. Although scoring criteria are not provided, examples are offered for some of the items. Since it is anticipated that the parent-teacher-child activities will take place in the home, it is recommended that a large portion of the assessment take place in the home as well.

Range of Children: The Oregon Project was developed for utilization with blind and visually handicapped young children ranging in age from birth to 6 years.

Testing Time: No information is reported.

Scoring: A Skills Inventory will be maintained for each child and may be used over successive years. A check (✓) is placed in the column to the right of the item if the child demonstrates proficiency with that particular skill. The date that the skill was assessed and found to be adequate is also recorded. No total score is obtained; therefore, no conversion to any kind of standard score is possible.

Examiner: The authors are not explicit on this point. However, in their statement of purpose, they state that the Skills Inventory has been developed to provide guidance to educators of young children with visual deficits.

Standardization: No information reported.

Reliability: No information reported.

Validity: No information is reported regarding research efforts to establish content, concurrent, or predictive validity.

The Oregon Project began as an attempt to revise the original Portage Guide to Early Education. The Skills Inventory is based upon records of preschoolers in the Southern Oregon Program for Visually Impaired, current literature in the field, and input from teachers of preschool blind children.

Discussion/Concerns: The Skills Inventory is organized much like many of the developmental profiles measuring the growth of young children. It does include many items, however, that are unique to the development of the visually handicapped child. The Inventory might have greater utility when used in conjunction with either the *Developmental Profile* (Alpern and Boll, 1972) or *A Social Maturity Scale for Blind Preschool Children* (Maxfield and Buchholz, 1957). It should provide users with information that is useful in the preparation of individualized educational plans and as a record of the child's growth and development over time.

A particularly helpful and interesting aspect related to curriculum planning is the author's discussion of each of the six developmental areas and their significance to the young visually handicapped child.

In summary, the Oregon Project provides a guide for educators serving visually handicapped children. The teaching activities are clearly stated and have been used successfully in home-based service delivery models. The authors report that the materials have also been used by itinerant teachers of visually impaired, regular classroom teachers who are mainstreaming young blind children and teachers in institutional classroom settings. It should have good utility as one of several resources used by those educating young visually handicapped children.

References:

Brown, D., Simmons, V. & Methvin, J. (1978). *The Oregon Project for Visually Impaired and Blind Preschool Children*. Medford, Oregon: Jackson County Education Service District.

Social Skills Rating System (SSRS)—Preschool Level

Author(s):	Frank M. Gresham & Stephen N. Elliott
Publisher:	American Guidance Service
Address:	Circle Pines, MN 55014-1796
Copyright Date:	1990
Price:	$27—manual $16.50—parent or teacher questionnaires (30)

Purpose: The SSRS can be used to screen and classify children suspected of having significant social behavior problems and helps in identifying appropriate interventions for identified children.

Description: The SSRS for preschoolers includes teacher and parent rating scales which assess two domains: Social Skills and Behavior Problems. A student form is available for older students. Dependent on which rating scales are completed, the SSRS Social Skills Scale includes five subdomains: cooperation, assertion, responsibility, empathy, and self-control. The SSRS Problem Behavior Scale looks at externalizing problems, internalizing problems, and hyperactivity. These scales can be integrated on a record form which attempts to link assessment results and intervention strategies.

Range of Children: The SSRS can be used to assess children from age 3 to 12th grade. The preschool level assess children from 3-0 to 4-11.

Testing Time: It takes approximately 20 minutes for one rating scale to be completed.

Scoring: The respondent fills out a questionnaire form which includes both frequency-of-behavior scoring and importance-of-behavior scoring. Raw scores can be converted into standard scores and percentile ranks. Scores can also be converted into Behavior Level scores which are discussed in the manual.

Examiner: Professional, usually a psychologist.

Standardization: The SSRS was standardized on a national sample of 4,170 children using their self-ratings along with ratings of the children made by 1,027 parents and 259 teachers. The SSRS was representative of sex and grade and included subjects from four U.S. regions, various community sizes, different races, special education students, and regular education students. Teachers were predominantly female and were from different community sizes. Parents included both fathers and mothers and were representative of the U.S. population in terms of race and educational level.

Reliability: Internal consistency of the SSRS is discussed in the manual. Median correlation coefficient for the Social Skills Scale was .90, while it was .84 for the Problem Behaviors Scale and .95 for the Academic Competence Scale. Test-retest reliability correlations were .85 for Social Skills, .84 for Problem Behaviors, and .93 for Academic Competence. Interrater reliability is briefly discussed in the manual. According to Gresham and Elliott (1990), "we expect different raters to contribute unique views of the subject being rated"; consequently, correlations were not expected to be high.

Validity: The authors of the SSRS thoroughly discuss content, social, criterion-related, construct, concurrent, convergent, and discriminant validity. According to Gresham and Elliott (1990), validity is well supported; however, low coefficients in some areas are of concern.

Discussion/Concerns: Strengths of the SSRS are as follows: (1) it emphasizes positive behaviors, (2) it incorporates a multi-rater approach, (3) the importance rating scale aids in developing intervention goals, (4) and the integrative record form helps to link assessment with intervention strategies. The standardization procedures appear adequate. Validity and reliability are well discussed in the manual; however, there is some concern as to their adequacy. Overall, the SSRS appears to be a welcome edition to rating scales and could be used as a screening instrument and/or as a supplement to other assessments.

References:

Gresham, F. M. & Elliott, S. N. (1990). *Social Skills Rating System.* Circle Pines, MN: American Guidance Services.

Stanford-Binet Intelligence Scale: Fourth Edition (SB:FE)

Author(s): R. L. Thorndike, E. P. Hagen, & J. M. Sattler
Publisher: The Riverside Publishing Company
Address: 8420 Bryn Mawr Ave., Chicago, IL 60631
Copyright Date: 1986
Price: $453.60

Purpose: The SB:FE is used to assess general intelligence.

Description: The SB:FE is an individually administered intelligence test. It consists of 15 subtests which are not used through all ages of the scale. Examinees are first given the Vocabulary subtest, which is considered a routing test, since performance on this subtest along with the child's age will determine the entry level for the rest of the subtests. Preschool children are typically administered the following subtests: Vocabulary, Comprehension, Absurdities, Memory for Sentences, Pattern Analysis, Copying, Quantitative, and Bead Memory. Two factor scores (Verbal Comprehension and Nonverbal Reasoning/Visualization) are calculated from these subtests.

Range of Children: The SB:FE can be used to assess individuals from 2-0 to adult.

Testing Time: Depending on how many subtests are administered, the SB:FE takes approximately 1½ to 2½ hours.

Scoring: Raw scores can be converted into three types of standard scores: standard age scores ($\underline{M}$ = 50, $\underline{SD}$ = 8), area scores ($\underline{M}$ = 100, $\underline{SD}$ = 16), and a Composite Score ($\underline{M}$ = 100, $\underline{SD}$ = 16). The SB:FE allows for the computation of area and Composite Scores when less than the entire battery is administered.

Examiner: SB:FE is administered by a psychologist.

Standardization: The standardization sample consisted of 5,013 individuals in 17 age groups. The number of individuals in each age group ranged from 194 to 460. The sample was representative of the U.S. population according to 1980 census data. Stratification variables included geographic region, community size, ethnic group, age, gender, and socioeconomic status. (Sattler, 1990)

Reliability: Internal consistency for the Composite Score ranged from .95 to .99 across the age ranges with a median reliability of .97. The median subtest reliabilities ranged from .73 to .94. Test-retest reliability was assessed using two age groups and a time period of two to eight months.

BEST COPY AVAILABLE

Composite score coefficients were .91 and .90. The subtest correlation coefficients were highly variable ranging from .56 to .78 for one age group and .28 to .86 for the other. Standard Error of Measurement in scaled score points is 2.8 for the Composite Score.

Validity:

The SB:FE has been compared to several other intelligence tests in order to assess concurrent validity. Correlation coefficients ranged from .27 to .91 with a median of .80 as reported in the manual. Researchers have also undertaken the task of assessing SB:FE's concurrent validity and have generally found that "for populations within the average intellectual range, the SB:FE is likely to yield Composite Scores that are similar to those provided by the WISC-R, WAIS-R, and Form L-M" (Sattler, 1990). Construct validity has been established in the following ways: (1) raw scores increase as a function of age, (2) factor analyses support a number of dimensions of the scale, and (3) subtests correlate moderately to highly with the Composite Score.

Discussion/Concerns:

Psychometric properties of the SB:FE are considered generally adequate as it has good validity, high reliability, and excellent standardization (Sattler, 1990). However, it is important to note that the adequacy of validity and reliability of subtest scores markedly decreases when interpreted alone. Limitations of the SB:FE are as follows: (1) lack of a comparable battery throughout the age ranges covered by the scale; (2) variable range of scores; (3) limited support for the four area scores; (4) difficulty in scoring some responses; (5) lack of description of procedure for establishing cutoff criteria; (6) inappropriate entry level points; and (7) overly long administration time (Sattler, 1990).

Strengths of the SB:FE are as follows: (1) technically adequate test, (2) good administration procedures, (3) adequate administrative guidelines and test materials, and (4) helpful scoring criteria (Sattler, 1990). The SB:FE is potentially a very effective assessment tool for assessing cognitive abilities and can be considered one of the premiere tests of intelligence.

References:

Sattler, J. M. (1990). *Assessment of Children (3rd Edition)*. San Diego, CA: Jerome M. Sattler.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Guide for Administering and Scoring the Fourth Edition*. Chicago: Riverside Publishing.

Structured Photographic Expressive Language Test—Preschool (SPELT-P)

Author(s):	Ellen O'Hara Werner & Janet Dawson Kresheck
Publisher:	Janelle Publications
Address:	Box 811, Dekalb, IL 60115-0811
Copyright Date:	1983
Price:	$59 per kit

Purpose: The SPELT-P identifies those children who may have difficulty with expression of early developing morphological and syntactic features. It is designed to be a screening instrument and does not diagnose the presence of language delays.

Description: The SPELT-P was designed in basically the same manner as the SPELT-II (Berryman, 1985). The examiner verbalizes the stimulus item and presents an accompanying photo to which the child responds. The photographs are colorful and realistic, and show people, animals, and objects in clearly defined situations. The administrator is allowed to use prompting when necessary to secure a response from the child. The test contains 25 items: 6 items sample early developmental structures and the other 19 have been simplified or modified from the SPELT-II (Berryman, 1985). On several items a sentence completion format is employed.

The SPELT-P also contains scoring criteria for African-American dialect but includes no norms. A Spanish form of the SPELT-P is also available.

Range of Children: The SPELT-P was designed to assess children from ages 3-0 to 5-11.

Testing Time: Test administration takes approximately 3 to 10 minutes depending on the child and the examiner's experience with giving this test.

Scoring: Scoring is basically the same as the procedure used for the SPELT-II. Responses are scored as correct, incorrect, or incorrect-no response. Raw scores can be compared to means, standard deviations, and cut-off scores.

Examiner: Professional, usually a speech and language clinician.

Standardization: The norm group for the SPELT-P was based on 731 Caucasian monolingual children who had no apparent handicapping conditions. The children were randomly selected from preschools in three states: Florida, Illinois, and Missouri. The norm group was generally representative of age, sex, and community size. The children were primarily from middle socioeconomic homes.

Reliability: Berryman (1985) states that "high test-retest and internal consistency reliability values are reported for all SPELT-P age levels." Test-retest reliability was assessed using a sample of 48 children and coefficients ranged from .86 to .97. Internal consistency coefficients ranged from .84 to .92 using all subjects in the standardization pool.

Validity: According to Berryman (1985), content validity is justified through extensive documentation; however, construct and concurrent validity are considered insufficient. For concurrent validity, the SPELT-P was compared with the Preschool Language Scale which resulted in a correlation coefficient of .76. The test is considered to effectively differentiate between normal children and language-delayed children; however, this conclusion is based on a sample of only 21 children. The test authors report that mean scores did increase with age.

Discussion/Concerns: The standardization procedures appear inadequate for the SPELT-P as the norm group is not representative. Information on validity and reliability is insufficient since small sample sizes were used. A strength of the SPELT-P is its appealing format which will be of interest to young children. Werner and Kresheck (1983) state that scores derived from the SPELT-P should be interpreted with caution because language development of young children is highly variable and can change dramatically. The SPELT-P is not a diagnostic assessment tool, but a screening test of language development.

References:

Berryman, J. D. (1985). Review of the Structured Photographic Expressive Language Test—II. In J. T. Mitchell (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Werner, E. O. & Kresheck, J. D. (1983). *The Structured Photographic Expressive Language Test—Preschool.* Sandwich, IL: Janelle Publications.

Structured Photographic Expressive Language Test (SPELT-II)

Author(s):	Ellen O'Hara Werner & Janet Dawson Kresheck
Publisher:	Janelle Publications, Inc.
Address:	Box 811, Dekalb, IL 60115-0811
Copyright Date:	1983
Price:	$69 per kit

Purpose: The purpose of the SPELT-II is to measure the generation of specific morphological and syntactic structures. It allows analysis of a child's ability to use several common grammatical forms, as well as to perform rule-governed changes in sentence structure.

Description: The SPELT-II attempts to assess a child's language level by eliciting responses from the child in a contextual setting through structured visual and auditory stimuli (Werner & Kresheck, 1983). The stimulus for each of the 50 items is a photograph accompanied by oral cueing. Administration instructions direct the user to employ appropriate prompting, short of supplying the target structure for immediate imitation, to gain a response. The photographic stimuli are colorful and realistic, and the response forms facilitate ease of administration (Berryman, 1985). A Spanish form of the SPELT-II is also available.

Range of Children: The SPELT-II is designed to test children from ages 4-0 to 9-5.

Testing Time: Testing takes approximately 15 to 25 minutes depending on the child and the examiner's experience with administering the SPELT-II.

Scoring: Responses are scored as correct, incorrect, or incorrect (no response). The child's score is the total number correct. Means and standard deviations, percentile ranks, and mean percent correct are provided for six month age levels spanning the age range. An alternative scoring procedure using African-American dialect is available; however, there are no available norms for this procedure (Berryman, 1985).

Examiner: Professional, usually a speech and language clinician.

Standardization: The normative population of 1,178 children was based on children who had no apparent developmental problems. The sample was randomly selected from schools in the North-Central and Southern regions of the United States. The children were primarily from middle socioeconomic homes and were from both urban and rural communities. The sample was representative of age and sex.

Reliability:

Test-retest and internal consistency coefficients are high at the younger age levels; however, ceiling effects of this test reduce reliability values at older age levels (Berryman, 1985). A test-retest reliability of .91 was derived using a sample of 81 children. In addition, a sampling of a group of language-delayed children yielded a test-retest coefficient of .87. Internal consistency ranged from .70 to .87.

Validity:

According to Berryman (1985), content validity is well documented and discussed in the manual; however, construct and concurrent validity are considered generally inadequate, due to weaknesses in how the data were gathered. Concurrent validity was assessed by comparing the SPELT-II with the Test of Language Development using only 20 children. An 86 percent agreement resulted. The manual states that a correlation coefficient of .97 between a child's age and score was found. Also, Werner and Kreshek indicate that the SPELT-II can effectively discriminate between normal and language-delayed children; however, a limited sample group was used to assess this.

Discussion/Concerns:

The psychometric properties of the SPELT-II appear somewhat inadequate, especially regarding validity and standardization procedures. Sample sizes were generally inadequate. The "ceiling effect" of this test may result in over-identification of older children (Berryman, 1985). Also, the norm group appears to be based on "normal" children which may also result in over-identification of children regardless of age. Scoring procedures include an incorrect-no response choice which may inappropriately depress a child's score due to a lack of understanding of the directions rather than a lack of knowledge. Berryman concluded that this test "provides a clinically useful measure of expressive language structure when employed with children from standard English speaking backgrounds and who possess normal or near normal oral language comprehension skills."

The authors state the SPELT-II does give some indications of a child's functional level in relationship to peers. However, they caution that this test gives only a general idea regarding a child's language performance. Consequently, use of this assessment tool for diagnostic purposes should be avoided.

References:

Berryman, J. D. (1985). Review of Structured Photographic Expressive Language Test-II. In J. T. Mitchell (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Werner, E. O. & Kresheck, J. D. (1983). *The Structured Photographic Expressive Language Test—II*. Sandwich, IL: Janelle Publications.

Test for Auditory Comprehension of Language—Revised Edition (TACL-R)

Author(s):	Elizabeth Carrow-Woolfolk
Publisher:	DLM Teaching Resources
Address:	One DLM Park, Allen, TX, 75002, 1-800-527-4747
Copyright Date:	1985
Price:	$130 per kit

Purpose:

As the manual states, this test provides a means of assessing a child's auditory comprehension, helps to identify individuals with receptive language disorders, helps to guide the clinician toward specific areas that need additional testing, and provides a means of measuring change in auditory comprehension. It specifically assesses comprehension of grammatical form and structure, content, and vocabulary.

Description:

The TACL-R is an individually administered test which includes three major sections. Section I assesses word classes and relations. Section II is called Grammatical Morphemes and includes short sentences and verbal stimuli which are designed to reflect semantic modulations of other structures such as tense or mood. Section III, Elaborated Sentences, assesses comprehension of more complex sentences of various types such as interrogatives and negatives. The test involves having the child point to one of three pictures which best describes the test items presented verbally by the examiner. One of the three pictures of each item depicts the meaning of a word, morpheme or syntactic structure while the other two pictures illustrate a decoy or contrast to the stimuli. The TALC-R includes more pictures of minorities than did the earlier version of the test. (Bankson, 1989).

Range of Children:

The TACL-R can be used to assess children ranging in age from 3-0 to 9-11.

Testing Time:

Average testing time is from 10 to 20 minutes.

Scoring:

Basal and ceiling scoring rules are provided and test items are ordered according to difficulty. Raw scores can be converted to percentile ranks, standard scores, and age equivalents. Non-normalized standard scores are also provided for use with extremely low or high scores.

Examiner:

Professional, usually speech clinician.

Standardization:

Normative data are provided for ages 3-0 to 9-11 and for grades kindergarten to four. The norming sample was representative of geographical regions in the U.S. as well as socioeconomic status, age, sex, and race. One thousand three children, 161 of whom were of ethnic/minority backgrounds, were included.

Handicapped children and hearing-impaired children were also tested to assess test adaptability (Bankson, 1989).

Reliability: Internal consistency was demonstrated with a split-half reliability of .96. Test-retest reliability ranges from .89 to .91 across age levels. Standard errors of measurement are reported for each age group, and are fairly low (Bankson, 1989).

Validity: Content validity is supported by the rationale and procedures involved in item selection which have been examined by numerous reviewers and test users (Bankson, 1989). Construct validity is adequately demonstrated, and concurrent validity is reflected in high correlations between scores on the TALC-R and several appropriate tests.

Discussion/Concerns: The TACL-R is a carefully and systematically constructed test with adequate reliability, validity, and standardization procedures (Bankson, 1989). It is fairly easy to administer and score. This instrument can be used with confidence when normative data are needed for a particular student. It can be highly useful as part of a comprehensive language evaluation. If used for program planning, the TALC-R does not provide adequate information concerning specific problems a student may be having. Overall, this test is a good measure of language comprehension within a very specific examination context (Bankson, 1989).

References:

Bankson, N. W. (1989). Review of the Test for Auditory Comprehension of Language—Revised Edition. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language—Revised*. Austin, TX: DLM Teaching Resources.

The Test of Early Language Development (TELD)

Author(s):	Wayne P. Hresko, D. Kim Reid, Donald D. Hammill
Publisher:	Western Psychological Services
Address:	12031 Wilshire Boulevard, Los Angeles, CA 90025
Copyright Date:	1981
Price:	$60

Purpose:

The TELD was developed to identify those children who are significantly behind their peers in the development of language, to document children's progress in language, to serve as a measure in research projects, and to suggest instructional practices.

Description:

The TELD is comprised of 38 items which assess language content and form. The form of language refers to syntax, morphology, and phonology. In TELD, syntax and morphology are measured both receptively and expressively. Phonology is measured only productively with the child's pronunciation of words. Emphasis is placed on the syntactic aspect of form because of its central role in the transmission of meaning. Language content refers to specific word knowledge, knowledge of conceptual categories, and interpretation of meaning within various contexts. The child's ability to express and receive meaning is evaluated.

Range of Children:

3-0 through 7-11.

Testing Time:

Average of 15-20 minutes.

Scoring:

A child's TELD test performance is reported in terms of three kinds of normative scores, Language Quotients (LQs), Percentiles, and Language Ages (LAs). The LQ is a standard score with a mean of 100 and a standard deviation of 15.

Examiner:

Professional, usually a speech clinician.

Standardization:

The TELD was standardized on the test performance of 1,184 children who live in 11 states and one Canadian province. Where possible, the sample's characteristics are compared with those of the United States' population as reported in the *Statistical Abstract of the United States (1979)*. Six geographic regions were sampled and race, community size, and socioeconomic status were considered.

Reliability:

Internal consistency coefficients ranged from .87 to .92 for five different age levels. Total test reliability was reported at .90 with a standard error of measurement of two (Dole, 1985). Test-retest reliability was found to be .90 for a time span of approximately two weeks (177 children were used in this study spanning five age groups).

Validity:

According to Dole (1985), content validity was adequately documented in terms of item selections, sampling, difficulty, and discrimination. Concurrent validity with the Test of Language Development was .80; however, a small sample was used. The TELD was also compared with a number of other assessments, but validity coefficients were found to be inadequate. Construct validity evidence is reported, but Dole concludes that it is difficult to evaluate.

Discussion/Concerns:

The authors note that caution should be used when interpreting the TELD scores since it does not measure all aspects for early language development. The manual provides detailed and explicit information on the theoretical foundation and interpretation of the test. Guidelines and precautions are very well documented (Prather, 1985). While the test's standardization and reliability aspects appear adequate, its validity is inadequate. Its emphasis on verbal expressive language and content is a strength (Prather). It is an extremely useful SCREENING tool of language development for children ages three to eight; however, it is not recommended for use as a diagnostic tool, and it provides limited information regarding program planning (Dole, 1985). The authors state that the TELD should be used along with other assessments.

References:

Hresko, W. P., Reid, D. K., & Hammill, D. D. (1981). *Administration Manual, Test of Early Language Development*. Los Angeles: Western Psychological Services.

Dole, J. A. (1985). Review of the Test of Early Language Development. In J. T. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Prather, E. M. (1985). Review of the Test of Early Language Development. In J. T. Mitchell, Jr. (Ed.), *The Ninth Mental Measurements Yearbook*. Lincoln: The University of Nebraska Press.

Test of Early Socioemotional Development (TOESD)

Author(s):	Wayne P. Hresko and Linda Brown
Publisher:	Pro-Ed
Address:	8700 Shoal Creek Boulevard, Austin, TX, 78758 (512) 451-3246
Copyright Date:	1984
Price:	$74 per complete kit

Purpose: To provide a measure of children's behavior across several settings such as school and home.

Description: The TOESD is composed of four independent components which are as follows: (1) student rating scale—30 items to which the student responds "yes" or "no", (2) Teacher rating scale of 36 descriptive phrases that are evaluated by the teacher using a four-point scale, (3) Parent rating scale of 34 items rated by parents or a caretaker using a four-point scale, and (4) sociogram—a peer nominating technique using positive choices (who would you invite to a party?).

Range of Children: 3-0 to 7-11 years.

Testing Time: Rating scales take about 20 minutes each to complete. The sociogram takes approximately 30 minutes.

Scoring: Rating scales are easy to score and take approximately 10 minutes, while the sociogram takes about 30 minutes. Results can be reported in percentile ranks and standard scores.

Examiner: The manual states the test may be administered by any professional who has read the manual.

Standardization: Procedures appear appropriate, and the norm group is based on the 1983 Statistical Abstract of the United States to include race, ethnicity, socioeconomic status, and geographic location. Minority groups are included; however, low socioeconomic status children may be underrepresented. The manual doesn't describe how the sample was selected (Doll, 1989).

Reliability: Internal consistency and test-retest reliability are adequate with standard error of measurement being low. No reliability information is provided for the sociogram since it is difficult to measure the reliability of this assessment tool.

Validity: Concurrent validity was assessed using three measures of preschool behavior: Basic School Skills Inventory (BSSI), Behavior Evaluation Scale (BES), and Behavior Rating Profile (BRP). The TOESD correlated adequately with these measures. However, the TOESD and BRP share numerous items which results in validity being spuriously high (Doll, 1989).

Furthermore, the BRP was normed on a higher age group and correlations cannot be generalized to a younger age group (Doll, 1989). Content validity and construct validity are not discussed in the manual.

Discussion/Concerns:

This model of multiple measures based on the same norm group is considered to be an effective assessment tool. Reviewers indicate reliability is adequate while validity is not considered adequate. Also, the purpose of this test is to assess socioemotional competencies; however, it appears to actually measure socioemotional deficits. Item wording tends to be quite negative. Lower scores on the TOESD are evidence of problematic behaviors; however, whether or not these deficits are equivalent in importance is not taken into account. Overall, this measure is not recommended for use alone, but could be used as a supplemental tool with other assessments.

References:

Doll, E. J. (1989). Review of the Test of Early Socioemotional Development. In J. Close Conoley & J. J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook*. Lincoln: University of Nebraska Press.

122

Test of Language Development—2 Primary (TOLD-2:P)

Author(s):	Phyllis L. Newcomer & Donald D. Hammill
Publisher:	Pro-Ed
Address:	8700 Shoal Creek Blvd., Austin, TX 78758
Copyright Date:	1988
Price:	$119 per complete kit

Purpose: The revised TOLD-2:P is designed to identify specific receptive and expressive language skills of primary age children. Results reveal strengths, weaknesses, and irregularities in specific areas of language development. It does not assess a child's ability to use language for communication or learning purposes (Westby, 1988).

Description: The TOLD-2:P Primary assesses semantic abilities by having the child define words and point to pictures that best represent words spoken by the examiner. Syntactic abilities are assessed by having the child: (1) repeat verbatim sentences of increasing difficulty, (2) complete sentences begun by the examiner, and (3) select one of three pictures that best represents a sentence stated by the examiner. Phonology is assessed based on the child's production when labeling pictures and the child's judgement of whether two words spoken by the examiner are the same or different. The test includes seven subtests: Oral Vocabulary, Picture Vocabulary, Grammatic Understanding, Sentence Imitation, Grammatic Completion, Word Articulation, and Word Discrimination. A detailed description of administration procedures is provided in the manual.

Range of Children: The TOLD-2:P can be administered to children from ages 4-0 to 8-11.

Testing Time: Administration time is approximately 40 minutes.

Scoring: Raw scores can be converted to standard scores, percentiles, age equivalents, and composite quotients. The composite quotients can be divided into the following: Spoken Language (SLQ), Listening (LiQ), and Semantics (SeQ).

Examiner: Professional, usually a speech and language clinician.

Standardization: The TOLD-2:P was standardized on over 2,436 children from 29 states and one province. The sample approximated the national population in 1985 in relation to sex, residence, race, geographic distribution, and occupation of parents.

Reliability: Internal consistency of the TOLD-2:P was assessed in several studies. Median correlation coefficients of .90 were obtained for the subtests and .94 for the composite scores in one study. Among a sample of children having oral communication problems, coefficients for the subtests ranged from .80 to .89; the coefficient of the composite score was .95.

Test-retest reliability coefficients ranged from .92 to .96 for composite scores. Subtest scores were somewhat lower. Standard Error of Measurements for the TOLD-2:P were fairly low. "Well established procedures were used to determine reliability related to internal test consistency and stability of test-retest scores" (Westby, 1988).

Validity:

Content validity was assessed by having professionals rate the degree to which the subtests measured the test dimensions to include listening, speaking, phonology, syntax, and semantics. Criterion-related validity was assessed by comparing the TOLD-2:P with several other language tests. Correlation coefficients were highly variable with many being in the .70 range. The TOLD-2:P and the Peabody Picture Vocabulary Test resulted in a coefficient of .80. Construct validity was purported to be established by studying the relationship of the TOLD-2:P to age, IQ, and school achievement; by factor analyzing scores; and by distinguishing between groups of children with language problems.

Discussion/Concerns:

According to Westby (1988), "the TOLD-2 Primary . . . [is] well designed in terms of meeting established psychometric criteria." The manual clearly discusses standardization procedures, validity, and reliability. In general, the TOLD-2:P is considered a screening test rather than a diagnostic instrument. According to the authors, results derived from the TOLD-2:P should be the first step in a comprehensive evaluation of a student's problem in language and should be followed with other assessment procedures. The TOLD-2:P is a widely used test and is considered to be a valuable resource for those who assess or teach in the area of spoken language.

References:

Newcomer, P. L. & Hammill, D. D. (1988). *Test of Language Development—Primary (2nd Ed.)*. Austin, TX: Pro-Ed.

Westby, C. (1988). Test Review: Test of Language Development—2 Primary. *The Reading Teacher, 41*, 236-237.

Uniform Performance Assessment System (UPAS)

Author(s):	Norris G. Haring, Owen R. White, Eugene B. Edgar, James Q. Affleck, & Alice H. Hayden
Publisher:	Psychological Corporation
Address:	7555 Caldwell Avenue, Chicago, IL 60648
Copyright Date:	1981
Price:	$95 for complete kit $12 for record forms (12)

Purpose: The UPAS is specifically designed to assist teachers and parents in identifying just how far a pupil has progressed in the development of the basic skills normally acquired during the first six years of life.

Description: The UPAS is criterion referenced and individually administered. Four separate curricular areas are assessed: (1) Pre-academic/Fine Motor Development, (2) Communication, (3) Social/Self-help Skills, and (4) Gross Motor Development. One additional section addresses problems dealing with specific inappropriate behaviors (e.g., tantruming or self-abuse). The UPAS consists of approximately 250 items, with between 45 and 76 items on any given curricular area. In order to administer the assessment, the UPAS Criterion Test manual, stimulus cards, and record forms are needed. The manual describes each item and how it should be assessed. The materials, test/observation procedure, and criteria for scoring each item are included on each Criteria Test card. A majority of the test materials need to be supplied by the examiner.

Range of Children: Birth to 6 years.

Testing Time: No administration time is reported

Scoring: Each item is scored + (pass) or - (fail). In addition, a coding system for procedures that are atypical or have had to be modified is included. Subtotals for each category are calculated and converted into percentage scores. These percentages are used with a progress chart to determine instructional goals and accomplished goals.

Examiner: Teachers and Parents

Standardization: No information is available

Reliability: Reliability seems to be adequate, although limited information is available. Two types of reliability were calculated: test-retest and procedural adequacy. Test-retest reliabilities are reported on 80 assessments, one week apart. Agreements ranged from 86 percent to 100 percent agreement between two different testers' scores at two different times. Procedural accuracy, an index of the degree to which observers agree

BEST COPY AVAILABLE

that a tester followed specified testing procedures, ranged from 94 percent to 100 percent accuracy.

Validity: Validity studies for the UPAS are inadequate. The authors claim that the UPAS has content validity due to the fact that the developers felt that it measured the curriculum they taught. In addition, the test has been used at numerous other facilities and, since no complaints of content inappropriateness have been received, the authors report that this suggests that the scale may be valid in other settings as well.

Discussion/Concerns: The UPAS differs from other testing instruments with regard to allowing certain items to be classified as "outcome category," thereby assessing the child's ability to accomplish activities regardless of the manner in which the information is expressed (Gresham, 1985). According to reviewers, the scale seems to assess adequately the skills of handicapped children. However, the technical information for the UPAS is inadequate and should be improved.

References:

Haring, N. G. (1981). *Uniform Performance Assessment System Criterion Tests*. Columbus, OH: Charles E. Merrill.

Gresham, F. (1985). Review of the Uniform Performance Assessment System. In J. V. Mitchell (Ed.), *The Ninth Mental Measurements Yearbook* (Vol II, pp. 1640-1641). Lincoln: University of Nebraska.

Vineland Adaptive Behavior Scales

Author(s):	Sara S. Sparrow, David A. Balla, and Domenic V. Cicchetti
Publisher:	American Guidance Service
Address:	Publishers Building, Circle Pines, MN 55014-1796
Copyright Date:	1984, 1985 (Classroom Edition) Interview Edition
Price:	$85.35 for (Survey & Expanded Forms) $31.25 per Classroom Edition $107.95 per Complete Scale Set

Purpose:

The Vineland is an adaptive behavior instrument that is intended to assess an individual's performance on the daily activities required for personal and social self-sufficiency. It can be used in a variety of educational, clinical, or research settings. Assessment of adaptive behavior functioning in the mental retardation population is emphasized as an important feature of the Vineland; however, it can be used as a diagnostic and/or educational planning instrument for a variety of populations.

Description:

The Vineland Scales are administered individually to a respondent (parent, caregiver, or classroom teacher) who is familiar with the daily activities of the individual being assessed. The Scales measure adaptive behavior in four domains: Communication, Daily Living Skills, Socialization, and Motor Skills. The Expanded and Survey Forms also include a maladaptive behavior domain. Each domain is further divided into subdomains.

There are three versions of the Vineland: The Interview Edition-Expanded Form, The Interview Edition-Survey Form, and the Classroom Edition. The Expanded Form includes 577 items. This form offers a more comprehensive assessment of adaptive behavior and a systematic basis for planning individual educational or treatment programs. The Survey From includes 297 items from the Expanded Form. The Classroom Edition includes 244 items and assesses behavior in the classroom.

Range of Children:

The Survey and Expanded Forms assess individuals ages birth to 18-11 and low functioning adults. The Classroom Edition is appropriate for students 3-12 years of age.

Testing Time:

Expanded Form: approximately 60-90 minutes

Survey Form:

approximately 20-60 minutes

Classroom Edition:

approximately 20 minutes

Scoring: The Vineland Scales have standard scores for the domains and the Adaptive Behavior Composite; each has a mean of 100 and a standard deviation of 15. Each version yields standard scores and age equivalents. Adaptive levels and age equivalents are given for each of the subdomains. On all forms, item scores reflect whether or not the individual performs the activity described. The raw scores are then converted to standard scores, percentile ranks, stanines, and age-equivalent scores based on the standardization sample. Bands of error are also computed. Conversion Tables are provided in the manual. The Adaptive Behavior Composite score is obtained by adding the subdomain scores, and standard scores are calculated from tables in the manual.

Examiner: Psychologist, social worker, or other professional with a graduate degree and specific training in individual assessment and test interpretation.

Standardization: The standardization sample for all three forms was representative of the population as described by 1980 census data. Stratification variables included sex, race or ethnic group, geographical region, community size, and parents' educational level. The Survey and Expanded Forms were based on a sample of 3,000 individuals. Norms are provided from birth to 18 years, 11 months. Separate norms also are available for mentally retarded, emotionally disturbed, and physically handicapped children and adults. The Classroom Edition had a representative sample of 2,984 students.

Reliability: The Vineland Scales generally have sufficient reliability. The manual reports three measures of reliability, although not for each edition. Test-retest reliability for the Survey Form ranges from the .80s to .90s. Test-retest reliability coefficients were not computed for the Expanded or Classroom Forms.

Split-half coefficients for the Survey Form were calculated for each domain. The median coefficients for the Survey Form ranged from a low of .83 (Motor Skills) to a high of .94 (Adaptive Behavior Composite). Split-half coefficients for the Expanded Form were estimated based on the Survey Form and adjusted by the Spearman-Brown formula. Median reliability coefficients ranged from a low of .91 (Motor Skills) to .97 (Adaptive Behavior Composite). Median split-half reliability coefficients for the Classroom Edition ranged from .80 (Motor Skills) to .98 (Behavior Composite). Full coefficient ranges for each domain are included in the manual.

Interrater reliability for the Survey and Expanded Forms ranged from .62 to .75. Standard errors of measurement and standard deviations are also reported in the manual.

Validity:

The validity of the Vineland Scales appears to be adequate. A major premise of the test is that adaptive behavior is age-related and, therefore, scores on adaptive behavior measures should increase with age. The authors report that construct validity of the Vineland is supported by data which demonstrate that the scores on the Vineland subdomains do rise progressively with age.

Content validity is reported by the authors as good for all forms. Field testing, national item tryout, and national standardization produced a set of items that assess adaptive behavior according to the definition of adaptive behavior adopted by the authors.

Concurrent validity was low to inadequate. It was established by correlating the Vineland Adaptive Behavior Scales with the AAMD Adaptive Behavior Scale and various other tests. Correlations between the Vineland and the AAMD were low.

Discussion/Concerns:

Overall, the Vineland's psychometric properties are adequate meaning that it is a potentially useful tool for the assessment of adaptive behavior. The manual provides detailed information regarding the technical data of the instrument. However, there are some concerns. Fluctuations in the means and standard deviations of the standard scores from age group to age group are a problem, especially with mentally retarded individuals. According to Sattler (1986), these fluctuations mean that critical differences needed to evaluate significant differences among domain scores may be only rough approximations and it may be difficult to compare individuals across ages. Difficulties, especially for inexperienced users in scoring responses, framing questions, and eliciting appropriate responses is also something of a concern.

Finally, a number of users report that age-equivalent scores vary widely across domains and subdomains even when standard scores do not. It is recommended, therefore, that the age scores be interpreted with extreme caution.

References:

Sattler, J. M. (1989). Review of Vineland Adaptive Behavior Scales. In J. Close-Conley & J. Kramer (Eds.), *Tenth Mental Measurements Yearbook* (Vol. I, pp. 878-881). Lincoln: University of Nebraska Press.

Sparrow, S. S., Balla, D. A., and Cicchetti, D. V. (1984). *Interview Edition-Expanded Form Manual-Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., and Cicchetti, D. V. (1984). *Interview Edition Survey Form Manual-Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Sparrow, S. S., Balla, D. A., and Cicchetti, D. V. (1985). *Classroom Edition-Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R)

Author(s):	David Wechsler
Publisher:	The Psychological Corporation
Address:	555 Academic Court, San Antonio, Texas 78204-2498, 1-800-228-0752
Copyright Date:	1989
Price:	$445 per complete set

Purpose: The WPPSI-R is a standardized measure of the intellectual abilities of young children. It measures intelligence by using various tasks. "A child's performance on these tasks is summarized by a composite score, which serves as an estimate of the individual's capacity to understand and cope with the surrounding world" (Wechsler, 1989).

Description: The WPPSI-R is an individually administered test which consists of 12 subtests: Object Assembly, Geometric Design, Block Design, Mazes, and Picture Completion in the Performance Scale; and Information, Comprehension, Arithmetic, Vocabulary, and Similarities comprise the Verbal Scale. There are two optional subtests: Animal Pegs (Performance Scale) and Sentences (Verbal Scale). Test items are presented by an examiner and require a variety of responses, some verbal and some nonverbal, from the child. "Essentially, the WPPSI-R can be considered a downward extension of the WISC-R" (Sattler, 1990).

Range of Children: The WPPSI-R can be administered to children from ages 3-0 to 7-3.

Testing Time: Administration takes approximately one hour and 15 minutes.

Scoring: Raw scores are converted to scaled scores and deviation IQ scores by age. Subtest scaled scores are expressed as standard scores ($\underline{M}$ = 10, $\underline{SD}$ = 3). The Full Scale IQ score has a mean of 100 and a standard deviation of 15. When fewer than 10 subtests are administered, IQs can be computed by prorating or by a special short-form procedure. Test-age-equivalent scores are also provided in the manual.

Examiner: Test is administered by psychologists.

Standardization: The WPPSI-R was standardized on 1,700 children. U.S. census data (1986) were used to select representative children for the normative sample. The sample was stratified by age, race, gender, geographic region, parents' education, and parents' occupation.

Reliability: Average internal consistency reliabilities are .92 for the Performance Scale, .95 for the Verbal Scale, and .96 for the Full Scale. Subtest reliability coefficients range from .63 to .86. Lowest reliabilities are found at age 7.

Test-retest reliability was studied using 175 children and a time period of three to seven weeks. Correlation coefficients for the Performance, Verbal, and Full Scales were .87, .89, and .91 respectively. Standard error of measurement in IQ points is 3.00 for the Full Scale.

Validity:

Concurrent validity was assessed comparing the WPPSI-R to several intelligence tests. Correlation coefficients suggest that the WPPSI-R yields IQs similar to a number of intelligence tests. Construct validity is supported through factor analysis and appears to be a fair measure of intelligence (Sattler, 1990). However, Sattler contends that additional research is needed using various age levels, especially the young age groups, as this is lacking.

Discussion/Concerns:

The WPPSI-R is a well-standardized test, with good reliability and validity. However, low reliability of some individual subtests is a concern. Limitations of the WPPSI-R include the following: (1) limited floor which does not differentiate abilities well, (2) nonuniformity of subtest scores, (3) long administration time, (4) possible difficulties in scoring responses, and (5) problems for some minority children and for children who do not work quickly (Sattler, 1990). It is also important to note that the WPPSI-R and WISC-III are not considered parallel forms. Strengths of this instrument are that it yields useful diagnostic information, it has good administration procedures, materials are colorful and of high interest, and the manual is easy to use (Sattler, 1990). Although the WPPSI-R has some limitations, it is, overall, a valuable tool for the assessment of young children's intelligence.

References:

Sattler, J. M. (1990). *Assessment of Children*. San Diego: Jerome M. Sattler.

Wechsler, D. (1989). *WPPSI-R Manual*. San Antonio, TX: Psychological Corporation.

Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R)

Author(s): Richard W. Woodcock & M. Bonner Johnson
Publisher: DLM Teaching Resources
Address: One DLM Park, Allen, TX 75002
Copyright Date: 1989
Price: $555 complete WJ-R; $235 - Achievement Battery; $370 - Cognitive Battery

Purpose: The WJ-R is a comprehensive diagnostic tool consisting of both cognitive and achievement tests which were co-normed.

Description: The WJ-R contains five tests for measuring early cognitive development in the Standard Cognitive Battery and six tests for measuring early achievement in the Standard Achievement Battery. The five cognitive tests are Memory for Names, Memory for Sentences, Incomplete Words, Visual Closure, and Picture Vocabulary. There are also four supplemental subtests which can provide additional information for preschoolers. The preschool achievement tests include Letter-Word Identification, Applied Problems, Dictation, Science, Social Studies, and Humanities. For the cognitive battery, the child is required to respond verbally to test items presented by the examiner or from a tape played by the examiner. The achievement battery involves written responses from the child. There are two forms (A and B) of the Achievement tests.

Range of Children: The WJ-R can be administered to anyone over 2 years of age. The Early Development Scale is typically administered to preschoolers.

Testing Time: Administration of the cognitive and achievement batteries take approximately 30-60 minutes each. It is recommended that administration be completed in two settings.

Scoring: During administration, responses are recorded onto a test record. Basal and ceilings levels are included as well as clear scoring directions. Raw scores can be converted to W scores, age equivalents, grade equivalents, relative mastery index scores, percentile ranks, and standard scores. Computer scoring is available and is highly recommended since hand scoring is quite complicated.

Examiner: Professional, usually a special educator or psychologist.

Standardization: Norms are based on a stratified random sample balanced in terms of the national distributions on sex, race, Hispanic origin, occupation, geographic location, and type of community. The norming sample included 6,359 subjects ranging in age from 24 months to 95 years. Subjects came from widely distributed communities throughout the U.S. The preschool sample consisted of 705 subjects.

Reliability: Internal consistencies for the Early Development Scale (cognitive) subtests ranged from .81 to .95, with the Broad Cognitive Ability coefficient being .95. Standard errors of measurement are fairly low. Internal consistencies for the Achievement scale subtests ranged from .81 to .95. No test-retest reliability was reported in manual.

Validity: The manual provides evidence of content validity through a discussion of test development. Concurrent validity was assessed at three different age levels and compared to several other tests measuring cognitive ability. The Early Development Scale was compared to several tests; correlation coefficients ranged from .48 (Boehm Tests of Basic Concepts) to .69 (Stanford Binet—Fourth Edition). For the Achievement battery, correlation coefficients were highly variable and fairly low when compared to other achievement tests. The highest correlations reached the .60 to .71 range for both the Knowledge and Skills factors. The manual indicates that construct validity is supportable as scores improve with age, average children can be distinguished from lower-functioning children, and tests appear to measure different aspects of cognitive abilities and achievement skills.

Discussion/Concerns: The WJ-R is a well-standardized test and has generally adequate validity and reliability. However, concurrent validity is considered inadequate, content validity is not well discussed in the manual, and there is no information regarding test-retest reliability of this instrument. Scoring of the WJ-R is cumbersome and time-consuming; however, computer scoring is available. Materials are two-dimensional, with no manipulables or tags at the level of younger children. Lower functioning preschoolers may not basal on this scale. A strength of the WJ-R is that it includes co-norms for the cognitive and achievement batteries. Also, given the fact that it spans a large age range, children can be repeatedly tested using this instrument. The manuals are easy to follow and contain good information concerning administration and scoring. Overall, the WJ-R is a valuable instrument for assessing cognitive and achievement abilities, but is limited in its usefulness with preschoolers suspected of developmental delays.

References:

Woodcock, R. W. & Mather, N. (1989). *Woodcock-Johnson Tests of Achievement: Standard and Supplemental Batteries*. Allen, TX: DLM Teaching Resources.

Woodcock, R. W. & Mather, N. (1989). *Woodcock-Johnson Tests of Cognitive Ability: Standard and Supplemental Batteries*. Allen, TX: DLM Teaching Resources.